Subset Semantics for Justifications

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Eveline Lehmann

aus Basel

Leiter der Arbeit:
Prof. Dr. Th. Studer
Institut für Informatik

Subset Semantics for Justifications

Inauguraldissertation
der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Eveline Lehmann

aus Basel

Leiter der Arbeit:
Prof. Dr. Th. Studer
Institut für Informatik

Von der Philosophisch-naturwissenschaftlichen Fakultät angenommen.

Bern, den 13. Januar 2021

Der Dekan:
Prof. Dr. Z. Balogh

Acknowledgements

A few years ago I had to decide whether or not to start this PhD. I have not been sure about it, since I was quite happy with the job I had. Obviously, I decided to embark on this adventure. Although it was not always easy, I have never regretted my decision. This is largely thanks to my supervisor Prof. Dr. Thomas Studer, who guided me with a sure hand through all the obstacles. I would like to thank him for all his support and guidance.

There are a lot of other people who supported me during this time. I would especially like to thank Prof. em. Dr. Gerhard Jäger, Bettina Choffat, Dr. Peppo Brambilla, and all the other members of the Logic and Theory group for their support, for the coffee breaks and all those nice little moments. Special thanks go to Atefeh Rohani for all the wonderful conversations and many sweet little things.

Outside of our institute I would like to thank Prof. Dr. George Metcalfe and his group as well as all my colleagues at Gymnasium Burgdorf and Gymnasium Lerbermatt.

In addition, I would like to thank all my friends and familiy for their patience and support. I would especially like to thank my godmother Patricia Schilt-Barth, not only for the opportunity to escape theory and paperwork by spending time in a vineyard, but especially for her lifelong support.

Finally, I would also like to thank Theo Oggier and Agnes Imhof for their valuable feedback on a preliminary version of this thesis and for their friendship and Florian Zierer for preparing the LaTeX-template of this thesis.

The research for this thesis was supported by the Swiss National Science Foundation.

Contents

Introduction

Justification logic is a variant of modal logic where the $\Box$-operator is replaced with justification terms that stand for specific justifications. So a formula $\Box A$ from modal logic which can be read as *there is **a** proof / **an** evidence for A* is replaced by $t : A$ with the intended meaning that ***the** justification t is a justification to belief A* . In justification logic we ask about the reason why we believe or know something [8, 9, 23].

The first justification logic was the Logic of Proofs and was developed by Artemov [1, 5]. The aim was to provide intuitionistic logic with classical provability semantics. In this logic the justification terms represent proofs in a formal system like Peano arithmetic. By *proof* we mean a Hilbert-style proof which is a sequence of formulas

$$F_1, \ldots F_n \tag{*}$$

where each formula is either an axiom or derived by a rule application from formulas that occur earlier in the proof. In this interpretation $t : A$ holds if A occurs in the proof-sequence represented by t. Hence t is not only a proof for the last formula in the sequence F_n but for all formulas in it.

The benefit of being able to represent explicit reasons in logical systems is not limited to mathematical proof systems but can be used to represent evidence in general. Using this interpretation, justification logic provides a versatile framework for epistemic logic [2, 6, 12, 16, 21]. If we interpret terms in this context, we ignore the order of the sequence in (*) and interpret evidence terms as sets of formulas.

This is anticipated in both Mkyrtychev models [28] as well as Fitting models [19]. The former are used to obtain a decision procedure for justification logic where one of the main steps is to keep track of which (set of) formulas a term justifies, see, e.g. [23, 34]. The latter provide first epistemic models for justification logic based on possible worlds semantics. Each world in the model has its own evidence function that specifies, which term serves as possible evidence for which (set of) formulas in this world.

Artemov [3] introduced modular models where justification terms are interpreted as sets of propositions when he addressed the problem of the logical type of justifications.

In this thesis justification terms are interpreted as sets of possible (or impossible) worlds instead. This is called a subset semantics since we define that $t : A$ holds if the interpretation of t is a subset of the extension of A. Various kinds of justification logics and the corresponding subset models will be presented.

The thesis consists of two main parts. The first part gives an overview of the general framework of subset models. The basic ideas are introduced and the standard logical axioms are discussed. In the second part the possibilities to use subset models are extended for other purposes. Besides standard extensions like introducing probabilities or dynamic aspects like updates, some new aspects like new operators for combining justifications and logics that deal with presumptions are introduced.

This thesis is based on four articles in collaboration with Thomas Studer: *Subset models for justification logic* [24], which is also submitted in an extended version for the journal *Information and Computation*, *Exploring subset models for justification logic* [26], which is submitted to be published in the collection with the working title *Research Trends in Contemporary Logic* by Melvin Fitting et al., *Belief Expansion in Subset Models* [25], which was presented at LNCS in 2020 and *Impossible and Conflicting Obligations in Justification Logic* [18], which will be presented at DEON 2021 and which is a collaboration with Faroldi and Ghari.

Part I.

Subset Models for Justification Logics

This Part introduces subset models for justification logics. We offer a new semantics to interpret justification terms so that they become more like pieces of evidence than just sets of formulas. In fact, this is not the first attempt to do so. Artemov [7] and Artemov and Nogina[10] treated justification terms as sets of possible worlds, a strategy which we follow too. However, they did not consider the usual term structure. In Chapter 6 we show that the subset models presented here can be used to model aggregating probabilistic evidence like proposed in [7]. On the other hand, there are several topological approaches to evidence available [11, 35, 36], which, however, do not feature justifications explicitly in their language.

In [33] Sedlár and Podroužek establish a relational semantics for justification logic J that looks quite similar to our subset semantics. However, they claim that the relation has to be irreflexive, which is a big obstacle to modelling aspects like factivity, as they mention themselves.

It is the aim of this part to provide a new semantics called *subset semantics* for justification logic that interprets terms as sets of possible worlds and operations on terms as operations on sets of possible worlds. We then say that $t : A$ is true if A is true in all the worlds belonging to the interpretation of t.

Usually, justification logic includes an application operator that represents modus ponens (MP) on the level of terms. We provide two approaches to handle this operator in our semantics. The first is to include a new constant $\mathsf{c}^\star$ which is interpreted as the set of all the worlds closed under (MP) and then use this new constant to define an application operator. We follow this approach in Chapter 1. The second way, which is investigated in Chapter 2, is to include an application operator directly. However, this leads to some quite cumbersome definitions. In Chapter 3 we investigate the differences between these two kinds of models according to monotonicity of application.

Another difference between our semantics and many other semantics for justification logic is that we allow non-normal (impossible) worlds. They are usually needed to model the fact that agents are not omniscient and that they do not see all the consequences of the facts they are already aware of. In Chapter 4 we show how these impossible worlds preserve the aspect of hyperintensionality, which is central to the logic of justification.[1]

[1] I would like to thank Igor Sedlár for his comments that have been very inspiring for a lot of aspects discussed in Chapter 3 and Chapter 4.

In the first two Chapters we present soundness and completeness proofs. However, completeness for logics that contain the axiom **jd** but not the axiom **jt** is only shown for logics, where all axioms are justified by some constant, so called axiomatically appropriate constant specifications. In Chapter 5 we finally demonstrate how our framework can be slightly modified to guarantee completeness for all constant specifications. Furthermore, we look at various representations of the D-axiom in modal logics and investigate the differences of their realizations in justification logics.

1. $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset models

In this chapter a first version of subset models for justification logics will be introduced. We do not limit ourselves to one logic, but we investigate a whole family of logics. As is usual in this field, we will introduce the language and the logics before we present the new semantics which will be proven to be sound and complete.

1.1. Syntax

Justification terms are built from countably many constants c_i and variables x_i and the special and unique constant $\mathsf{c}^\star$ according to the following grammar:

$$t ::= c_i \mid x_i \mid \mathsf{c}^\star \mid (t + t) \mid !t$$

The set of terms is denoted by Tm. The set of atomic terms, i.e. terms that do not contain any operator $+$ or $!$ is denoted by ATm. The operation $+$ is left-associative.

Formulas are built from countably many atomic propositions p_i, terms t and the symbol $\bot$ according to the following grammar:

$$F ::= p_i \mid \ \bot \ \mid F \to F \mid t : F$$

The set of atomic propositions is denoted by Prop and the set of all formulas is denoted by $\mathcal{L}_J$. The other classical Boolean connectives $\neg, \top, \wedge, \vee, \leftrightarrow$ are defined as usual.

Definition 1.1 ($\mathsf{c}^\star$**-term**)**.** A $\mathsf{c}^\star$**-term** is defined inductively as follows:

- $\mathsf{c}^\star$ is a $\mathsf{c}^\star$**-term**
- if s and t are terms and c is a $\mathsf{c}^\star$**-term** then $s + c$ and $c + t$ are $\mathsf{c}^\star$**-terms**

So a $\mathsf{c}^\star$**-term** is either $\mathsf{c}^\star$ itself or a sum-term where $\mathsf{c}^\star$ occurs at least once.

We investigate a family of justification logics that differ in their axioms and how the axioms are justified. We have two sets of axioms, the first axioms are:

cl	all axioms of classical propositional logic;
jc*	$c : A \wedge c : (A \to B) \to c : B$ for all $\mathsf{c}^\star$**-terms** c.
j+	$s : A \vee t : A \to (s+t) : A$;

The set of these axioms is denoted by $\mathsf{L}^\star_\alpha$.
There is another set of axioms:

j4	$t : A \to !t : (t : A)$;
jd	$t : \bot \to \bot$;
jt	$t : A \to A$.

This set is denoted by $\mathsf{L}^\star_\beta$. It is easy to see that **jd** is a special case of **jt**. By $\mathsf{L}^\star$ we denote all logics that are composed from the whole set $\mathsf{L}^\star_\alpha$ and some subset of $\mathsf{L}^\star_\beta$. Moreover, a justification logic $\mathsf{L}^\star$ is defined by the set of axioms and its constant specification CS that determines which constant justifies which axiom. So the constant specification is a set

$$\mathsf{CS} \subseteq \{(c, A) \mid c \text{ is a constant and } A \text{ is an axiom of } \mathsf{L}^\star\}$$

In this sense $\mathsf{L}^\star_{\mathsf{CS}}$ denotes the logic $\mathsf{L}^\star$ with the constant specification CS. To deduce formulas in $\mathsf{L}^\star_{\mathsf{CS}}$ we use a Hilbert system given by $\mathsf{L}^\star$ and the rules modus ponens:

$$\frac{A \qquad A \to B}{B}\ (\mathbf{MP})$$

and axiom necessitation

$$\frac{}{\underbrace{!\ldots!}_{n} c : \underbrace{!\ldots!}_{n-1} c : \ \ldots : !!c : !c : c : A}\ (\mathbf{AN!}) \quad \forall n \in \mathbb{N}, \text{ where } (c, A) \in \mathsf{CS}$$

1.2. Semantics

Definition 1.2 ($\mathsf{L}^\star_{\mathsf{CS}}$-subset models)**.** Given some logic $\mathsf{L}^\star$ and some constant specification CS, then an $\mathsf{L}^\star_{\mathsf{CS}}$-subset model $\mathcal{M} = (W, W_0, V, E)$ is defined by:

- W is a set of objects called worlds.
- $W_0 \subseteq W$ and $W_0 \neq \emptyset$.
- $V : W \times \mathcal{L}_J \to \{0,1\}$ such that for all $\omega \in W_0$, $t \in \mathsf{Tm}$, $F, G \in \mathcal{L}_J$:
  - $V(\omega, \perp) = 0$;
  - $V(\omega, F \to G) = 1 \quad \text{iff} \quad V(\omega, F) = 0 \text{ or } V(\omega, G) = 1$;
  - $V(\omega, t : F) = 1 \quad \text{iff} \quad E(\omega, t) \subseteq \{\, \upsilon \in W \mid V(\upsilon, F) = 1 \,\}$.
- $E : W \times \mathsf{Tm} \to \mathcal{P}(W)$ that meets the following conditions where we use
$$[A] := \{\omega \in W \;\mid\; V(\omega, A) = 1\}. \tag{1.1}$$
  For all $\omega \in W_0$, and for all $s, t \in \mathsf{Tm}$:
  - $E(\omega, s + t) \subseteq E(\omega, s) \cap E(\omega, t)$;
  - $E(\omega, \mathsf{c}^\star) \subseteq W_{MP}$ where W_{MP} is the set of deductively closed worlds, see below;
  - if **jd** $\in \mathsf{L}^\star$, then $\exists \upsilon \in W_0$ with $\upsilon \in E(\omega, t)$;
  - if **jt** $\in \mathsf{L}^\star$, then $\omega \in E(\omega, t)$;
  - if **j4** $\in \mathsf{L}^\star$, then
$$E(\omega, !t) \subseteq \{\, \upsilon \in W \mid \forall F \in \mathcal{L}_J \; (V(\omega, t : F) = 1 \Rightarrow V(\upsilon, t : F) = 1)\,\};$$
  - for all $n \in \mathbb{N}$ and for all $(c, A) \in \mathsf{CS}$: $E(\omega, c) \subseteq [A]$ and
$$E(\omega, \underbrace{!...!}_{n}\, c) \subseteq [\underbrace{!...!}_{n-1}\, c : ....!c : c : A].$$

The set W_{MP} is formally defined as follows:

$$W_{MP} := \{\omega \in W \mid \forall A, B \in \mathcal{L}_J \ ((V(\omega, A) = 1 \text{ and } V(\omega, A \to B) = 1) \\ \text{implies } V(\omega, B) = 1)\}.$$

So W_{MP} collects all the worlds where the valuation function is closed under modus ponens. W_0 is the set of *normal* worlds. The set $W \setminus W_0$ consists of the *non-normal* worlds. In a *non-normal* world both A and $\neg A$ may be true or none of them. Such *non-normal* or impossible worlds have been investigated by Veikko Rantala [30, 31] and in Chapter 4 we will show why they are necessary to keep the hyperintensional aspects of justification logic in subset models.

Moreover, using the notation introduced by (1.1), we can read the condition on V for justification terms $t : F$ as:

$$V(\omega, t : F) = 1 \quad \text{iff} \quad E(\omega, t) \subseteq [F]$$

Since the valuation function V is defined on worlds and formulas, the definition of truth is pretty simple:

Definition 1.3 (Truth in $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset models)**.** Let $\mathcal{M} = (W, W_0, V, E)$ be an $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset model, $\omega \in W$ and $F \in \mathcal{L}_J$. We define the relation $\Vdash$ as follows:

$$\mathcal{M}, \omega \Vdash F \quad \text{iff} \quad V(\omega, F) = 1$$

Remark 1.4. With the conditions on $E(w, \mathsf{c}^{\star})$ and $E(w, s + t)$ we obtain the intended meaning of a $\mathsf{c}^{\star}$**-term** $s + \mathsf{c}^{\star}$, namely that we consider only deductively closed worlds of s. However, the set $E(s + \mathsf{c}^{\star})$ does not have to be exactly the intersection of $E(w, s)$ with W_{MP} since we only have a subset-relation instead of equality. Hence $E(w, s + \mathsf{c}^{\star}) \neq E(w, \mathsf{c}^{\star} + s)$ in general. So even if in two $\mathsf{c}^{\star}$**-terms** the exactly same evidence sets occur, their order still matters. For the same reason $s + t : A \to t + s : A$ is not valid for any two distinct terms s and t.

1.3. Soundness

Definition 1.5 ($\mathsf{L}^{\star}_{\mathsf{CS}}$-validity)**.** Let CS be a constant specification. We say that a formula $F \in \mathcal{L}_J$ is $\mathsf{L}^{\star}_{\mathsf{CS}}$-*valid* if for each $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset model $\mathcal{M} = (W, W_0, V, E)$ and each $\omega \in W_0$, we have $\mathcal{M}, \omega \Vdash F$.

Since in non-normal worlds even the axioms of classical logic will not be valid in general, we only have soundness within W_0.

Theorem 1.6 (Soundness of $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset models). *For any justification logic $\mathsf{L}^{\star}_{\mathsf{CS}}$ and any formula $F \in \mathsf{L}^{\star}_{\mathsf{CS}}$ we have that if $\mathsf{L}^{\star}_{\mathsf{CS}} \vdash F$, then F is $\mathsf{L}^{\star}_{\mathsf{CS}}$-valid.*

Proof. The proof is by induction on the length of the derivation of F:

- If F is an instance of some axiom of classical logic, then the truth of F only depends on the valuation functions within the worlds of W_0. And all worlds of W_0 behave appropriately by definition.

- If F is derived by modus ponens, there is a $G \in \mathcal{L}_J$ s.t. $\mathsf{L}^{\star}_{\mathsf{CS}} \vdash G \to F$ and $\mathsf{L}^{\star}_{\mathsf{CS}} \vdash G$. By induction hypothesis $\mathcal{M}, \omega \Vdash G \to F$ hence we have $V(\omega, G \to F) = 1$ and therefore since $\omega \in W_0, \quad V(\omega, G) = 0$ or $V(\omega, F) = 1$ and again by induction hypothesis $\mathcal{M}, \omega \Vdash G$ and thus $V(\omega, G) = 1$. Together with $\omega \in W_0$, we obtain $V(\omega, F) = 1$, which is $\mathcal{M}, \omega \Vdash F$.

- If F is derived by axiom necessitation, then $F = c : A$ for some $(c, A) \in \mathsf{CS}$. By the condition on E within $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset models we have $E(\omega, c) \subseteq [A]$ for all $\omega \in W_0$. Hence $V(\omega, c : A) = 1$ and therefore $\mathcal{M}, \omega \Vdash c : A$. If F is a more complex formula like $!c : (c : A)$ derived by axiom necessitation, the argument is analogue.

- If F is an instance of the **jc***-axiom, then

  $$F = c : A \wedge c : (A \to B) \to c : B$$

  for some $A, B \in \mathcal{L}_J$ and a $\mathsf{c}^{\star}$**-term** c.
  Suppose that $\mathcal{M}, \omega \Vdash c : A$ and $\mathcal{M}, \omega \Vdash c : (A \to B)$. i.e. $E(\omega, c) \subseteq [A]$ and $E(\omega, c) \subseteq [A \to B]$. Hence for all $\upsilon \in E(\omega, c)$ we obtain $V(\upsilon, A) = 1$ and $V(\upsilon, A \to B) = 1$. From the definition of $\mathsf{c}^{\star}$**-terms**, the conditions on $E(w, \mathsf{c}^{\star})$ and $E(w, s + t)$ for some terms s, t, we infer that $E(w, c) \subseteq W_{MP}$ and we conclude $V(\upsilon, B) = 1$ and hence $E(\omega, c) \subseteq [B]$ and this means that $\mathcal{M}, \omega \Vdash c : B$.

- If F is an instance of the $\mathbf{j+}$-axiom, then $F = s : A \vee t : A \to s + t : A$ for some $s, t \in \mathsf{Tm}$ and $A \in \mathcal{L}_J$.
  Suppose wlog. $\mathcal{M}, \omega \Vdash s : A$, by Definition 1.3 we get $V(\omega, s : A) = 1$

and by Definition 1.2 and the conditions on V for worlds in W_0, $E(\omega, s) \subseteq [A]$. Since $E(\omega, s+t) \subseteq E(\omega, s) \cap E(\omega, t) \subseteq E(\omega, s)$ we obtain that $E(\omega, s+t) \subseteq [A]$ and by the condition on E in W_0 in Definition 1.2 that $V(\omega, s+t : A) = 1$. Hence by Definition 1.3 $\mathcal{M}, \omega \Vdash s+t : A$.

- If F is an instance of the **jd**-axiom, then $F = t :\perp \rightarrow \perp$ for some $t \in \mathsf{Tm}$.
  Suppose towards a contradiction that $\mathcal{M}, \omega \Vdash t :\perp$ for some $t \in \mathsf{Tm}$, then by Definition 1.3 we obtain that $V(\omega, t :\perp) = 1$ and hence by the condition of E in the worlds of W_0, $E(\omega, t) \subseteq [\perp]$. Since $\mathcal{M}$ must be a **jd**-$\mathsf{L}^\star_{\mathsf{CS}}$-subset model we claim that $\exists \upsilon \in W_0$ s.t. $\upsilon \in E(\omega, t)$. From $\upsilon \in E(\omega, t)$ we derive by the condition on V in Definition 1.2 $\upsilon \in [\perp]$ or in other words $\upsilon \in (\upsilon' \in W \mid V(\upsilon', \perp) = 1)$ and hence $V(\upsilon, \perp) = 1$ and this contradicts the claim that $\upsilon \in W_0$.

- If F is an instance of the **jt**-axiom, then $F = t : A \rightarrow A$ for some $A \in \mathcal{L}_J$ and some $t \in \mathsf{Tm}$.
  Suppose $\mathcal{M}, \omega \Vdash t : A$. By Definition 1.3 there is $V(\omega, t : A) = 1$. By the condition on worlds in W_0 in Definition 1.2 we get $E(\omega, t) \subseteq [A]$. Since $\mathcal{M}$ is a **jt**-$\mathsf{L}^\star_{\mathsf{CS}}$-subset model, $\omega \in E(\omega, t)$ and therefore we conclude $\omega \in [A]$. Hence $V(\omega, A) = 1$ and by Definition 1.3 we obtain that $\mathcal{M}, \omega \Vdash A$.

- If F is an instance of the **j4**-axiom, then $F = t : A \rightarrow !t : (t : A)$ for some $A \in \mathcal{L}_J$ and $t \in \mathsf{Tm}$
  Suppose $\mathcal{M}, \omega \Vdash t : A$, then by Definition 1.3 there is $V(\omega, t : A) = 1$. By the condition on E for **j4**-$\mathsf{L}^\star_{\mathsf{CS}}$-subset models for all $\upsilon \in E(\omega, !t)$ we obtain $V(\upsilon, t : A) = 1$. Therefore $E(\omega, !t) \subseteq [t : A]$ and by Definition 1.2 there is $V(\omega, !t : (t : A)) = 1$ and again by Definition 1.3 we conclude $\mathcal{M}, \omega \Vdash !t : (t : A)$. □

The **j**-axiom $s : (A \rightarrow B) \rightarrow (t : A \rightarrow s \cdot t : B)$ is not part of our logic. Using the $(\mathsf{c}^\star)$-axiom, we can define an application operation such that the **j**-axiom is valid.

Definition 1.7 (Application)**.** We introduce a new abbreviation $\cdot$ on terms by:

$$s \cdot t := s + t + \mathsf{c}^\star$$

Lemma 1.8 (The "j-axiom" follows). *For all* $\mathcal{M} = (W, W_0, V, E)$, $\omega \in W_0$, $A, B \in \mathcal{L}_J$ *and* $s, t \in \mathsf{Tm}$*:*

$$\mathcal{M}, \omega \Vdash s : (A \to B) \to (t : A \to s \cdot t : B)$$

Proof. Suppose $\mathcal{M}, \omega \Vdash s : (A \to B)$ and $\mathcal{M}, \omega \Vdash t : A$. Therefore $E(\omega, s) \subseteq [A \to B]$ and $E(\omega, t) \subseteq [A]$. We find

$$E(\omega, s \cdot t) = E(\omega, s + t + \mathsf{c}^\star) \subseteq \\ E(\omega, s) \cap E(\omega, t) \cap E(\omega, \mathsf{c}^\star) \subseteq [A \to B] \cap [A] \cap E(\omega, \mathsf{c}^\star).$$

Hence for all $\upsilon \in E(\omega, s \cdot t)$ we have $V(\upsilon, A \to B) = 1$ and $V(\upsilon, A) = 1$ and $\upsilon \in E(\omega, \mathsf{c}^\star)$ and therefore $V(\upsilon, B) = 1$. Hence $E(\omega, s \cdot t) \subseteq [B]$ and we obtain $\mathcal{M}, \omega \Vdash s \cdot t : B$. □

Of course there is as well a derivation within any of the presented logics. We use CR as an abbreviation for classical reasoning.

$s : (A \to B) \to s + t : (A \to B)$	**j+**
$s + t : (A \to B) \to s + t + \mathsf{c}^\star : (A \to B)$	**j+**
$s : (A \to B) \to s + t + \mathsf{c}^\star : (A \to B)$	CR
$t : A \to s + t : A$	**j+**
$s + t : A \to s + t + \mathsf{c}^\star : A$	**j+**
$t : A \to s + t + \mathsf{c}^\star : A$	CR
$s + t + \mathsf{c}^\star : (A \to B) \to (s + t + \mathsf{c}^\star : A \to s + t + \mathsf{c}^\star : B)$	**jc***
$s : (A \to B) \to (t : A \to s + t + \mathsf{c}^\star : B)$	CR

1.4. Completeness

To prove completeness we will construct a canonical model and then show that for every formula F that is not derivable in $\mathsf{L}^\star_{\mathsf{CS}}$, there is a model $\mathcal{M}^C$ with a world $\Gamma \in W_0^C$ s.t. $\mathcal{M}^C, \Gamma \Vdash \neg F$. Like in the case of other semantics for justification logics, the completeness of logics containing (**jd**) is only given, if the corresponding constant specification is axiomatically appropriate. Before we start with the definition of the canonical model, we must do some preliminary work. We will first prove that our logics are conservative extensions of classical logic. With this result we can argue, that the empty set is consistent and hence can be extended to so-called maximal $\mathsf{L}^{\star}_{\mathsf{CS}}$-consistent sets of formulas. These sets will be used to build the W_0-worlds in the canonical model.

Theorem 1.9 (Conservativity). *All logics $\mathsf{L}^{\star}$ presented are conservative extensions of the classical logic CL, i.e. for any formula $F \in \mathsf{L_{cp}}$:*

$$\mathsf{L}^{\star} \vdash F \quad \Leftrightarrow \quad \mathsf{CL} \vdash F$$

Proof. Since $\mathsf{L}^{\star}$ is an extension of CL the right-to-left direction is obvious. To prove the direction from left to right we use a translation $\mathfrak{t} : \mathcal{L}_J \to \mathsf{L_{cp}}$:

$$\begin{aligned}
\mathfrak{t}(P) &:= P \\
\mathfrak{t}(\bot) &:= \bot \\
\mathfrak{t}(A \to B) &:= \mathfrak{t}(A) \to \mathfrak{t}(B) \\
\mathfrak{t}(s : A) &:= \mathfrak{t}(A)
\end{aligned}$$

This translation removes all justification terms from a given formula. Now we show by induction on the length of the derivation for some formula A that $\mathsf{CL} \vdash \mathfrak{t}(A)$ whenever $\mathsf{L}^{\star} \vdash A$ and note that $\mathfrak{t}(A) = A$ for any $A \in \mathsf{L_{cp}}$. The cases where A is an axiom of CL is then obvious, since all logics $\mathsf{L}^{\star}$ contain all axioms of CL.

- **cl**: If A is an instance of some axiom scheme in $\mathcal{L}_J$, then $\mathfrak{t}(A) = A$ is an instance of the same axiom scheme in CL.
- **jc***: $\mathfrak{t}(c : A \wedge c : (A \to B) \to c : B) = A \wedge (A \to B) \to B$, which is a classical tautology.
- **j+**: $\mathfrak{t}(s : A \vee t : A \to (s + t) : A) = A \vee A \to A$, which is a classical tautology.
- **j4**,**jd**,**jt**: All translations have the form $A \to A$, which is a classical tautology.
- **modus ponens**: If A is derived by modus ponens, then there is a formula B s.t. $\mathsf{L}^{\star} \vdash B \to A$ and $\mathsf{L}^{\star} \vdash B$ and by induction hypothesis $\mathsf{L_{cp}} \vdash \mathfrak{t}(B) \to \mathfrak{t}(A)$ and $\mathsf{L_{cp}} \vdash \mathfrak{t}(B)$ and hence $\mathfrak{t}(A)$ can be derived in CL by modus ponens.

- **axiom necessitation**: If A is derived by axiom necessitation, then A is of the form $c : B$ for some axiom B. But $\mathfrak{t}(c : B) = B$ and B is an axiom. $\square$

Definition 1.10 (Consistency)**.** A logical theory L is called consistent, if $\mathsf{L} \nvdash \perp$. A set of formulas $\Gamma \subset \mathcal{L}_J$ is called L-consistent if $\mathsf{L} \nvdash \bigwedge \Sigma \to \perp$ for every finite $\Sigma \subseteq \Gamma$. A set of formulas Γ is called maximal L-consistent, if it is L-consistent and none of its proper supersets is.

Since all presented logics are conservative extensions of CL , we have the following consistency result.

Lemma 1.11 (Consistency of the logics). *All presented logics are consistent.*

As usual, we have a Lindenbaum lemma and the usual properties of maximal consistent sets hold, see, e.g., [23].

Lemma 1.12 (Lindenbaum Lemma). *Given some logic* L*, then for each* L*- consistent set of formulas* $\Gamma \subset \mathcal{L}_J$ *there exists a maximal consistent set* Γ' *such that* $\Gamma \subseteq \Gamma'$*.*

Definition 1.13 (Canonical Model)**.** For a given logic $\mathsf{L}^{\star}_{\mathsf{CS}}$ we define the canonical model $\mathcal{M}^C = (W^C, W^C_0, V^C, E^C)$ by:

- $W^C = \mathcal{P}(\mathcal{L}_J)$.
- $W^C_0 = \left\{ \Gamma \in W^C \;\middle|\; \Gamma \text{ is maximal } \mathsf{L}^{\star}_{\mathsf{CS}} - \text{consistent set of formulas} \right\}$.
- $V^C : V^C(\Gamma, F) = 1 \quad \text{iff} \quad F \in \Gamma;$
- E^C : With $\Gamma/t := \{F \in \mathcal{L}_J \mid t : F \in \Gamma\}$ and

$$W^C_{MP} := \{\Gamma \in W^C | \forall A, B \in \mathcal{L}_J : \\ \text{if } A \to B \in \Gamma \text{ and } A \in \Gamma \text{ then } B \in \Gamma\}$$

  we define :

$$E^C(\Gamma, t) = \left\{ \Delta \in W^C_{MP} \;\middle|\; \Delta \supseteq \Gamma/t \right\} \text{ if } t \text{ is a } \mathsf{c}^{\star}\textbf{-term}$$
$$E^C(\Gamma, t) = \left\{ \Delta \in W^C \;\middle|\; \Delta \supseteq \Gamma/t \right\} \text{ otherwise.}$$

Not all models that correspond to Definition 1.13 are $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset models. It depends on the presence of axiom **jd** and the constant specification.

Definition 1.14 (axiomatically appropriate CS)**.** A constant specification CS is called *axiomatically appropriate* if for each axiom A, there is a constant c with $(c, A) \in \mathsf{CS}$.

Axiomatically appropriate constant specifications are important as they provide a form of necessitation [9].

Now we must show that the canonical model is in general an $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset model.

Lemma 1.15. *The canonical model $\mathcal{M}^C$ is an $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset model if either*

(1) $(\boldsymbol{jd}) \notin \mathsf{L}^{\star}_{\mathsf{CS}}$ *or*

(2) the constant specification CS *is axiomatically appropriate or* $(\boldsymbol{jt}) \in \mathsf{L}^{\star}_{\mathsf{CS}}$.

Proof. In order to prove this, we have to show that $\mathcal{M}^C$ meets all the conditions we made for the valuation and evidence function and the constant specification i.e.:

(1) $W_0^C \neq \emptyset$.

(2) For all $\Gamma \in W_0^C$:
   a) $V^C(\Gamma, \bot) = 0$;
   b) $V^C(\Gamma, F \to G) = 1 \quad \text{iff} \quad V^C(\Gamma, F) = 0$ or $V^C(\Gamma, G) = 1$;
   c) $V^C(\Gamma, t : F) = 1 \quad \text{iff} \quad E(\Gamma, t) \subseteq [F]$.

(3) For all $\Gamma \in W_0^C, F \in \mathcal{L}_J, s, t \in \mathsf{Tm}$:
   a) $E^C(\Gamma, s + t) \subseteq E^C(\Gamma, s) \cap E^C(\Gamma, t)$;
   b) $E^C(\Gamma, \mathsf{c}^{\star}) \subseteq W_{MP}^C$;
   c) If **jd** in $\mathsf{L}^{\star}$, then $\exists \Delta \in W_0^C$ with $\Delta \in E^C(\Gamma, t)$;
   d) If **jt** in $\mathsf{L}^{\star}$: $\forall \Gamma \in W_0^C$ and $\forall t \in \mathsf{Tm} : \Gamma \in E^C(\Gamma, t)$;
   e) If **j4** in $\mathsf{L}^{\star}$:

$$E^C(\Gamma, !t) \subseteq \{\Delta \in W^C | \forall F \in \mathcal{L}_J \\ (V^C(\Gamma, t : F) = 1 \Rightarrow V^C(\Delta, t : F) = 1)\};$$

f) For all $(c, A) \in \mathsf{CS}$: $E^C(\Gamma, c) \subseteq [A]$ and

$$E^C(\Gamma, \underbrace{!...!}_{n} c) \subseteq [\underbrace{!...!}_{n-1} c : ....!c : c : A] \text{ for all } n \in \mathbb{N}.$$

So the proofs are here:

(1) Since the empty set is proven to be $\mathsf{L}^\star_{\mathsf{CS}}$-consistent (see Lemma 1.11) it can be extended by the Lindenbaum Lemma to a maximal $\mathsf{L}^\star_{\mathsf{CS}}$-consistent set of formulas Γ with $\Gamma \in W_0^C$.

(2) Suppose $\Gamma \in W_0^C$:

a) We claim $V^C(\Gamma, \bot) = 0$: Suppose the opposite, this means that $V^C(\Gamma, \bot) = 1$ hence by the definition of V^C follows that $\bot \in \Gamma$. But this is a contradiction to the fact that Γ is consistent.

b) From left to right: Suppose $V^C(\Gamma, F \to G) = 1$, then by the definition of V^C, $F \to G \in \Gamma$. Since Γ is maximal $\mathsf{L}^\star_{\mathsf{CS}}$-consistent this implies by maximal consistency of Γ that $F \notin \Gamma$ or $G \in \Gamma$. Hence again by the definition of V^C, $V^C(\Gamma, F) = 0$ or $V^C(\Gamma, G) = 1$.

From right to left: Suppose $V^C(\Gamma, F) = 0$ or $V^C(\Gamma, G) = 1$, then by the definition of V^C either $F \notin \Gamma$ or $G \in \Gamma$. Since $\Gamma \in W_0^C, \Gamma$ is maximal $\mathsf{L}^\star$-consistent and hence in both cases by maximal consistency that $F \to G \in \Gamma$. But this means again by the definition of V^C that $V(\Gamma, F \to G) = 1$.

c) From left to right: Suppose $V^C(\Gamma, t : F) = 1$, then by Definition 1.13 $t : F \in \Gamma$. Hence with the definition of Γ/t we obtain $F \in \Gamma/t$. So for each $\Delta \in E^C(\Gamma, t), F \in \Delta$ (again by Definition 1.13). Hence for these Δ it follows by the definition of V^C that $V^C(\Delta, F) = 1$ and therefore $\Delta \in [F]$. Since this is true for all $\Delta \in E^C(\Gamma, t)$ we obtain $E^C(\Gamma, t) \subseteq [F]$.

From right to left: The proof is by contraposition.
Suppose $V^C(\Gamma, t : F) \neq 1$, then by definition of V^C $t : F \notin \Gamma$. We define a world Δ by $\Delta := \Gamma/t$. Since $\Delta \in \mathcal{P}(\mathcal{L}_J)$ we can be sure that Δ exists, i.e. $\Delta \in W$. Since $t : F \notin \Gamma$ it follows that $F \notin \Gamma/t$ and therefore $F \notin \Delta$. But obviously $\Delta \supseteq \Gamma/t$ hence $\Delta \in E^C(\Gamma, t)$. So we conclude $E^C(\Gamma, t) \nsubseteq [F]$.
It remains to show that in case of $t = \mathsf{c}^\star, \Delta := \Gamma/t \in W_{MP}^C$

since otherwise $\Delta \notin E^C(\Gamma, \mathsf{c}^\star)$. In fact this is the case. Since $\Gamma \in W_0^C$ we obtain that Γ is a maximal $\mathsf{L}^{\star}_{\mathsf{CS}}$-consistent set of formulas and hence, whenever $\mathsf{c}^\star : A, \mathsf{c}^\star : (A \to B) \in \Gamma$ then by **jc*** we obtain $\mathsf{c}^\star : B \in \Gamma$. This means that whenever $A \in \Delta$ and $A \to B \in \Delta$ then $B \in \Delta$. Hence $\Delta = \Gamma/\mathsf{c}^\star$ is closed under modus ponens and therefore $\Delta \in W_{MP}^C$. So together with the former reasoning $\Delta \in E(\Gamma, \mathsf{c}^\star)$.

(3) Suppose $\Gamma \in W_0^C$:

a) Given some $F \in \mathcal{L}_J, s, t \in \mathsf{Tm}$: We start by an observation on the relation between the sets $\Gamma/(s+t)$ and Γ/s for $\Gamma \in W_0^C$. If $s : A \in \Gamma$ then since Γ is maximal $\mathsf{L}^{\star}_{\mathsf{CS}}$-consistent $s + t : A \in \Gamma$ and therefore $\Gamma/s \subseteq \Gamma/(s+t)$. With the same reasoning we obtain $\Gamma/t \subseteq \Gamma/(s+t)$. Thus if $\Delta \supseteq \Gamma/(s+t)$ then $\Delta \supseteq \Gamma/s$ and $\Delta \supseteq \Gamma/t$. This means that $E^C(\Gamma, s+t) \subseteq E^C(\Gamma, s)$ and $E^C(\Gamma, s+t) \subseteq E^C(\Gamma, t)$.[2] Therefore

$$E^C(\Gamma, s+t) \subseteq E^C(\Gamma, s) \cap E^C(\Gamma, t).$$

b) This follows directly from the definition of $E^C(\Gamma, \mathsf{c}^\star)$.

c) If **jd** in $\mathsf{L}^\star$, either CS is axiomatically approporiate or (**jt**) $\in \mathsf{L}^\star$ too.

- CS is axiomatically appropriate.

  For any $\Gamma \in W_0^C$ we obtain $\neg(t :\perp) \in \Gamma$. Hence $\perp \notin \Gamma/t$. Suppose towards a contradiction that Γ/t is not $\mathsf{L}^{\star}_{\mathsf{CS}}$-consistent, i.e. there exist $A_1, \dots A_n \in \Gamma/t$ s.t.

  $$A_1, \dots, A_n \vdash_{\mathsf{L}^{\star}_{\mathsf{CS}}} \perp . \tag{1.2}$$

  This together with the construction of Γ/t leads to $t : A_1, \dots, t : A_n \in \Gamma$. Since CS is axiomatically appropriate we can use (1.2) to infer $t : A_1, \dots, t : A_n \vdash_{\mathsf{L}^{\star}_{\mathsf{CS}}} s(t) :\perp$, for some term $s(t)$ only based on t. Since Γ is assumed to

[2] Please note if $s = \mathsf{c}^\star$ or $t = \mathsf{c}^\star$ this only holds due to $E^C(\Gamma, s+t)$ being constrained in this case to W_{MP}. Special thanks for drawing my attention to this point goes to the anonymous referee who found a mistake in the original completeness proof of $\mathsf{L}^{\star}_{\mathsf{CS}}$.

be maximally consistent we can use (**jd**) and apply modus ponens to infer $\bot \in \Gamma$ which contradicts the assumption that Γ is consistent. Therefore Γ/t is $\mathsf{L}^\star_{\mathsf{CS}}$-consistent and can be expanded by the Lindenbaum Lemma to a maximal $\mathsf{L}^\star_{\mathsf{CS}}$-consistent set $\Delta \supseteq \Gamma/t$ with $\Delta \in W_0^C$ and $\Delta \in E^C(\Gamma, t)$.

- (**jt**) $\in \mathsf{L}^\star$:

  The claim is a direct consequence of property (3d) (see next item).

d) Assume for some $F \in \mathcal{L}_J, \Gamma \in W_0^C$, $t \in \mathsf{Tm}$ that $F \in \Gamma/t$, i.e. $t : F \in \Gamma$, since Γ is maximal $\mathsf{L}^\star_{\mathsf{CS}}$-consistent and $t : F \to F$ is an instance of the **jt**-axiom, we conclude that $F \in \Gamma$. Since F was arbitrary we obtain $\Gamma \supseteq \Gamma/t$ and hence $\Gamma \in E^C(\Gamma, t)$.

e) Suppose for some $\Delta \in E^C(\Gamma, !t)$, hence $\Delta \supseteq \Gamma/!t$. Then assume for some arbitrary $F \in \mathcal{L}_J$, $V(\Gamma, t : F) = 1$ i.e. by Definition 1.13 we obtain $t : F \in \Gamma$. Since Γ is maximal $\mathsf{L}^\star_{\mathsf{CS}}$-consistent and $t : F \to !t : (t : F)$ is an instance of the **j4**-axiom we obtain $!t : (t : F) \in \Gamma$ and hence $t : F \in \Gamma/!t$. But then $t : F \in \Delta$ and by Definition 1.13 it follows that $V^C(\Delta, t : F) = 1$. Since F was an arbitrary formula and Δ an arbitrary world of $E^C(\Gamma, !t)$ we conclude that the condition holds.

f) Suppose $(c, A) \in \mathsf{CS}$, maximal $\mathsf{L}^\star_{\mathsf{CS}}$-consistency implies for all $\Gamma \in W_0^C$ that $c : A \in \Gamma$. Hence $A \in \Gamma/c$ and for all $\Delta \in E^C(\Gamma, c)$ we obtain $A \in \Delta$ and therefore $E^C(\Gamma, c) \subseteq [A]$.

   Furthermore maximal $\mathsf{L}^\star_{\mathsf{CS}}$-consistency implies for all $\Gamma \in W_0^C$ by axiom necessitation that

$$\underbrace{!...!}_{n} c : ... :!c : c : A \in \Gamma.$$

   Hence

$$\underbrace{!...!}_{n-1} c : ... :!c : c : A \in \Gamma / \underbrace{!...!}_{n} c$$

and for all $\Delta \in E^C(\Gamma, \underbrace{!...!}_{n} c)$ we obtain

$$\underbrace{!...!}_{n-1} c : ... :!c : c : A \in \Delta$$

and therefore

$$E^C(\Gamma, \underbrace{!...!}_{n} c) \subseteq [\underbrace{!...!}_{n-1} c : ... :!c : c : A]. \qquad \square$$

The Truth Lemma follows very closely:

Lemma 1.16 (Truth Lemma). *Let $\mathcal{M}^C = (W^C, W_0^C, E^C, V^C)$ be a canonical model, then for any $\Gamma \in W_0^C$:*

$$\mathcal{M}^C, \Gamma \Vdash F \text{ if and only if } F \in \Gamma.$$

Proof.

$$\mathcal{M}^C, \Gamma \Vdash F \overset{\text{Def. 1.3}}{\Longleftrightarrow} V^C(\Gamma, F) = 1 \overset{\text{Def. 1.13}}{\Longleftrightarrow} F \in \Gamma. \qquad \square$$

Hence each maximal $\mathsf{L}^\star_{\mathsf{CS}}$-consistent set is represented by some world in the canonical model and thus completeness follows directly:

Theorem 1.17 (Completeness). *Given some logic $\mathsf{L}^\star$ and let CS be constant specification which is required to be axiomatically appropriate in case **jd** $\in \mathsf{L}^\star$. For each formula F we have that if F is $\mathsf{L}^\star_{\mathsf{CS}}$-valid, then $\mathsf{L}^\star_{\mathsf{CS}} \vdash F$.*

Proof. The proof works with contraposition: Assume that $\mathsf{L}^\star_{\mathsf{CS}} \nvdash F$. Then $\{\neg F\}$ is $\mathsf{L}^\star_{\mathsf{CS}}$-consistent and by the Lindenbaum Lemma contained in some maximal $\mathsf{L}^\star_{\mathsf{CS}}$-consistent world Γ of the canonical model $\mathcal{M}^C$. And then $\mathcal{M}^C, \Gamma \nVdash F$. $\square$

In Chapter 5 we will show, that subset models for (**jd**) can be adapted such that completeness holds for arbitrary constant specifications.

2. $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-subset models

In this part we present an alternative definition of subset models for justification logic that directly interprets the application operator. Hence, we work with the standard language of justification logic and we consider the **j**-axiom instead of the axiom (**jc***). The structure of this Chapter corresponds to that of Chapter 1.

2.1. Syntax

In this section, justification terms are built from constants c_i and variables x_i according to the following grammar:

$$t ::= c_i \mid x_i \mid (t \cdot t) \mid (t + t) \mid !t$$

This set of terms is denoted by Tm^{A}. The operations $\cdot$ and $+$ are left-associative and ! binds stronger than anything else. Formulas are built from atomic propositions p_i and the following grammar:

$$F ::= p_i \mid \perp \mid F \to F \mid t : F$$

The set of atomic propositions is denoted by Prop and the set of all formulas is denoted by $\mathcal{L}^{\mathsf{A}}_J$. Again we use the other logical connectives as abbreviations.

As in Chapter 1, we investigate again a whole family of logics. They are arranged in two sets of axioms. The first set, denoted by $\mathsf{L}^{\mathsf{A}}_\alpha$ contains the following axioms:

cl	all axioms of classical propositional logic;
j	$s : (A \to B) \to (t : A \to s \cdot t : B)$;
j+	$s : A \vee t : A \to (s + t) : A$.

The other is identical to $\mathsf{L}_{\beta}^{\star}$ (modulo the different language) and contains:

j4 $t : A \rightarrow !t : (t : A)$;
jd $t : \perp \rightarrow \perp$;
jt $t : A \rightarrow A$.

For the sake of uniformity we denote this set of axioms by $\mathsf{L}_{\beta}^{\mathsf{A}}$.

By L^{A} we denote all logics that are composed from the whole set $\mathsf{L}_{\alpha}^{\mathsf{A}}$ and some subset of $\mathsf{L}_{\beta}^{\mathsf{A}}$.

There are no differences between these logics and the ones of the former section except in case of application. Therefore we skip all the details already mentioned and proved before.

CS and $\mathsf{L}_{\mathsf{CS}}^{\mathsf{A}}$ are defined as before except that the corresponding logic has changed as mentioned. And deducing formulas in $\mathsf{L}_{\mathsf{CS}}^{\mathsf{A}}$ works the same as in the previous section.

2.2. Semantics

Definition 2.1 ($\mathsf{L}_{\mathsf{CS}}^{\mathsf{A}}$-subset models)**.** Given some logic $\mathsf{L}_{\mathsf{CS}}^{\mathsf{A}}$ then an $\mathsf{L}_{\mathsf{CS}}^{\mathsf{A}}$-subset model $\mathcal{M} = (W, W_0, V, E)$ is defined like an $\mathsf{L}_{\mathsf{CS}}^{\star}$-subset model where

$$E : W \times \mathsf{Tm}^{\mathsf{A}} \rightarrow \mathcal{P}(W)$$

meets the following condition for terms of the form $s \cdot t$:

$$E(\omega, s \cdot t) \subseteq \mathfrak{W}_w(s, t),$$

where we use

$$\mathfrak{W}_w(s, t) := \{v \in W \mid \forall F \in \mathsf{APP}_w(s, t)(v \in [F])\}$$

with

$$\begin{aligned}\mathsf{APP}_w(s, t) := \{F \in \mathcal{L}_J^{\mathsf{A}} \mid \exists H \in \mathcal{L}_J^{\mathsf{A}} \text{ s.t.}& \\ E(w, s) \subseteq [H \rightarrow F] \text{ and } E(w, t) \subseteq [H]\}.&\end{aligned}$$

The set $\mathsf{APP}_\omega(s,t)$ contains all formulas that are colloquially said derivable by applying modus ponens to a formula justified by s and a formula justified by t.

Truth in an $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-subset models is defined as before.

Definition 2.2 (Truth in $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-subset models)**.** For an $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-subset model $\mathcal{M} = (W, W_0, V, E)$ and a world $\omega \in W$ and a formula F we define the relation $\Vdash$ as follows:

$$\mathcal{M}, \omega \Vdash F \quad \text{iff} \quad V(\omega, F) = 1.$$

2.3. Soundness

Definition 2.3 ($\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-validity)**.** Let CS be a constant specification. We say that a formula $F \in \mathcal{L}^{\mathsf{A}}_J$ is $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-*valid* if for each $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-subset model $\mathcal{M} = (W, W_0, V, E)$ and each $\omega \in W_0$, we have $\mathcal{M}, \omega \Vdash F$.

Theorem 2.4 (Soundness of $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-subset models). *For any justification logic $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$ and any formula $F \in \mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$ we have that if $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}} \vdash F$, then F is $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-valid.*

Proof. The proof is by induction on the length of the derivation of F and it is analogue to the proof of Theorem 1.6. The only thing that changes is the case, where F is an instance of the **j**-axiom:
Then $F = s : (A \to B) \to (t : A \to s \cdot t : B)$ for some $s, t \in \mathsf{Tm}^{\mathsf{A}}$ and $A, B \in \mathcal{L}^{\mathsf{A}}_J$. Assume for some $\omega \in W_0$ that $\mathcal{M}, \omega \Vdash s : (A \to B)$ and $\mathcal{M}, \omega \Vdash t : A$. Then by Definition 2.2 $V(\omega, s : (A \to B)) = 1$. Since $\omega \in W_0$ we obtain $E(\omega, s) \subseteq [A \to B]$ and by the same reason $V(\omega, t : A) = 1$ and $E(\omega, t) \subseteq [A]$. From the definition of $\mathsf{APP}_\omega(r, s)$ we conclude that $B \in \mathsf{APP}_\omega(s, t)$. So for all $\upsilon \in E(\omega, s \cdot t)$ we obtain by the requirements of E that $V(\upsilon, B) = 1$ hence $E(\omega, s \cdot t) \subseteq [B]$. From this, the fact that $\omega \in W_0$ and the requirements of V in W_0 we obtain $V(\omega, s \cdot t : B) = 1$, which is by Definition 2.2 $\mathcal{M}, \omega \Vdash s \cdot t : B$. $\square$

2.4. Completeness

Before we start defining a canonical model, we have to do the same preliminary work for $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$ as we had to do in the previous section for $\mathsf{L}^{\star}_{\mathsf{CS}}$. Since the logics $\mathsf{L}^{\star}_{\mathsf{CS}}$ from the former section differ only in one axiom, i.e. **j** replaces **jc***, we skip all the parts that are already done and focus on the changes that it brings about.

As before, we have a conservativity and consistency result.

Theorem 2.5 (Conservativity). *All logics* L^{A} *presented are conservative extensions of the classical logic* CL*, i.e. for any formula* $F \in \mathsf{L}_{\mathsf{cp}}$*:*

$$\mathsf{L}^{\mathsf{A}} \vdash F \quad \Leftrightarrow \quad \mathsf{CL} \vdash F.$$

Lemma 2.6 (Consistency of L^{A}). *All logics in* L^{A} *are consistent.*

All the other ingredients we needed in the former section to define and further develop the canonical model were generally defined and proven and can be adopted without additional effort.

To prove completeness we define a canonical model as follows:

Definition 2.7 (Canonical Model)**.** For a given logic L^{A} and a constant specification CS we define the canonical model $\mathcal{M}^C = (W^C, W^C_0, V^C, E^C)$ by:

- $W^C = \mathcal{P}(\mathcal{L}^{\mathsf{A}}_J)$;
- $W^C_0 = \{\Gamma \in W^C \mid \Gamma \text{ is maximal } \mathsf{L}^{\mathsf{A}}_{\mathsf{CS}} - \text{consistent set of formulas}\}$;
- $V^C : V^C(\Gamma, F) = 1 \quad \text{iff} \quad F \in \Gamma$;
- $E^C : E^C(\Gamma, t) = \{\Delta \in W \mid \Delta \supseteq \Gamma/t\}$.

Now we must show that such a canonical model is in fact a subset model.

Lemma 2.8. *The canonical model* $\mathcal{M}^C$ *is an* $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$*-subset model if* L^{A}

- *does not contain (**jd**) or*
- *contains it but the corresponding constant specification* CS *is axiomatically appropriate.*

Proof. In order to prove that, we have to proceed in the same way as in the previous section, i.e. showing that $\mathcal{M}^C$ meets all the conditions we made for the valuation and the evidence function as well as the constant specification.

Since the canonical model is defined in the same way as the one of $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset models, the corresponding proofs can be reused (see Lemma 1.15). Nevertheless, we need to look at the differences. Instead of showing that $E^C(\Gamma, \mathsf{c}^\star) \subseteq W^C_{MP}$ we have to show that

$$E^C(\Gamma, s \cdot t) \subseteq \{\Delta \in W^C \mid \forall F \in \mathsf{APP}_\Gamma(s,t)(\Delta \in [F])\}.$$

Assume that we are given $\Gamma \in W^C_0$, $F \in \mathcal{L}^{\mathsf{A}}_J$, $s, t \in \mathsf{Tm}^{\mathsf{A}}$. Take any $\Delta \in E^C(\Gamma, s \cdot t)$, i.e. $\Delta \supseteq \Gamma/(s \cdot t)$. Hence for all F s.t. $s \cdot t : F \in \Gamma$ we know that $F \in \Delta$. Hence by the definition of V^C, we have $V(\Delta, F) = 1$ and therefore $\Delta \in [F]$.

It remains to show: if $F \in \mathsf{APP}_\Gamma(s,t)$ then $s \cdot t : F \in \Gamma$. Suppose for some formula F that $F \in \mathsf{APP}_\Gamma(s,t)$ then by definition of $\mathsf{APP}_\Gamma(s,t)$ we know that there is a formula H s.t. $E^C(\Gamma, s) \subseteq [H \to F]$ and $E^C(\Gamma, t) \subseteq [H]$. By using Lemma 2.8 (the part that corresponds to Lemma 1.15 (2c) we conclude $V^C(\Gamma, s : (H \to F)) = 1$ and $V^C(\Gamma, t : H) = 1$. Hence by the definition of V^C we obtain $s : (H \to F) \in \Gamma$ and $t : H \in \Gamma$ and since Γ is maximal $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-consistent and $s : (H \to F) \to (t : H \to s \cdot t : F)$ is an instance of the j-axiom we conclude that $s \cdot t : F \in \Gamma$. ☐

Lemma 2.9 (Truth Lemma). *Let $\mathcal{M}^C = (W^C, W^C_0, E^C, V^C)$ be some canonical $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-subset model, then for all $\Gamma \in W_0$:*

$$\mathcal{M}^C, \Gamma \Vdash F \text{ if and only if } F \in \Gamma.$$

Proof.

$$\mathcal{M}^C, \Gamma \Vdash F \overset{\text{Def. 2.2}}{\Longleftrightarrow} V^C(\Gamma, F) = 1 \overset{\text{Def. 2.7}}{\Longleftrightarrow} F \in \Gamma.$$ ☐

Theorem 2.10 (Completeness). *Given some logic L^{A} and let CS be constant specification which is axiomatically appropriate in case $\boldsymbol{jd} \in \mathsf{L}^{\mathsf{A}}$. For each formula F we have that if F is $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-valid, then $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}} \Vdash F$.*

Proof. The proof is analogue to the one of Theorem 1.17. ☐

3. Comparing $\mathsf{L}^{\star}_{\mathsf{CS}}$ and $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$

In the previous Chapters we introduced two different kinds of logics and semantics. Both of them use a subset relation to model justification. This leads to the question in which sense they differ.

Lemma 3.1 (Monotonicity of application in $\mathsf{L}^{\star}_{\mathsf{CS}}$). *In $\mathsf{L}^{\star}_{\mathsf{CS}}$ the application operator is monotone, i.e.*

$$s : A \to s \cdot t : A, \text{ for all } s, t \in \mathsf{Tm}, A \in \mathcal{L}_J$$

Proof. This follows directly from axiom **j+** and Definition 1.7:

$$
\begin{array}{llr}
\text{(axiom } \mathbf{j+}\text{)} & s : A \to s + t : A & (3.1) \\
\text{(axiom } \mathbf{j+}\text{)} & s + t : A \to s + t + \mathsf{c}^{\star} : A & (3.2) \\
\text{(Definition 1.7)} & s + t + \mathsf{c}^{\star} : A = s \cdot t : A & (3.3) \\
\text{(3.1, 3.2, 3.3 and logical reasoning)} & s : A \to s \cdot t : A & \square
\end{array}
$$

In the corresponding semantics this fact holds because if $E(\omega, s) \subseteq [A]$, then any intersection of $E(\omega, s)$ with some other set will be as well a subset of $[A]$.

This phenomenon illustrates the intended meaning of $\cdot$ in $\mathsf{c}^{\star}$-subset models: if we have a justification s which justifies some formula A and we consider other justifications and have the capacity to apply modus ponens, this so combined justification will justify not less than s alone.

Lemma 3.2 (Monotonicity of application on $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$). *In $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$ application is not monotone.*

Proof. The proof is with a counterexample and by using soundness.

Consider the $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-subset model $\mathcal{M} = (W, W_0, V, E)$ with $W = \{ \omega_1, \omega_2 \}$, $W_0 = \{ \omega_1 \}$, $V(\omega_1, A) = 1$ and for all other formulas the valuation in w_1 is arbitrary but such that the conditions for V in W_0 are fulfilled, $V(\omega_2, X) = 0$ for all $X \in \mathcal{L}_J$, moreover $E(\omega_1, s) = \{ \omega_1 \}$, $E(\omega_1, t) = \{ \omega_2 \}$,

$E(\omega_1, s \cdot t) = \{ \omega_2 \}$ and all other justifications are defined, s.t. they fulfil the conditions of E in worlds of W_0.

First we have to prove that $\mathcal{M}$ is an $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-subset model, i.e. to show that all conditions made on V and E in Definition 2.1 for any $\omega \in W_0$ are fulfilled. For V this holds by definition of V in ω_1. So it remains to show that E behaves properly. Again, for all justifications except $E(\omega_1, s \cdot t)$ this holds by definition. So let us check whether

$$E(\omega_1, s \cdot t) \subseteq \mathfrak{W}_{\omega_1}(s, t). \tag{3.4}$$

We start by investigating the set $\mathsf{APP}_{\omega_1}(s,t)$: Since $E(\omega_1, s) = \{\omega_1\}$ and $V(\omega_1, A) = 1$ and $\omega_1 \in W_0$, we obtain in a first step that $\{\omega_1\} \subseteq [X \to A]$ for all $X \in \mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$ and as a consequence $V(\omega_1, s : (X \to A)) = 1$ for all those formulas X. Moreover $E(\omega_1, t) = \{\omega_2\}$ and since in ω_2 all formulas are evaluated to 0, we obtain that t justifies nothing. So there is no X with $V(\omega_1, t : X) = 1$. Hence $\mathsf{APP}_{\omega_1}(s,t) = \emptyset$.

Since $\mathsf{APP}_{\omega_1}(s,t) = \emptyset$, we obtain that $\mathfrak{W}_{\omega_1}(s,t) = W$. Therefore (3.4) is obvious. So $\mathcal{M}$ is indeed an $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-subset model.

Further we have $\mathcal{M}, \omega_1 \Vdash s : A$ but since $V(\omega_2, A) = 0$, there is

$$\{ \omega_2 \} = E(\omega_1, s \cdot t) \not\subseteq [A]$$

and hence $\mathcal{M}, \omega \nVdash s \cdot t : A$.

Finally, with Theorem 2.4 we conclude that $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}} \nvdash s : A \to s \cdot t : A$ $\square$

So the meaning of the application operator in $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$ is different to the one in $\mathsf{L}^{\star}_{\mathsf{CS}}$. In $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$, a term of the form $s \cdot t$ only justifies formulas that can be obtained by modus ponens; whereas in $\mathsf{L}^{\star}_{\mathsf{CS}}$, the term $s \cdot t$ justifies all formulas that are justified at least by one of its subterms s or t.

Another difference between the application in $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset models and models for standard justification logic is that application does ignore which justification justifies the condition and which justifies the antecedent.

Lemma 3.3. *For all* $\mathsf{L}^{\star}_{\mathsf{CS}}$*-subset models* $\mathcal{M} = (W, W_0, V, E)$*,* $\omega \in W_0$*,* $A, B \in \mathcal{L}_J$ *and* $s, t \in \mathsf{Tm}$*:*

$$\mathcal{M}, \omega \Vdash s : (A \to B) \to (t : A \to t \cdot s : B)$$

Proof. The proof is analogue to the one of Lemma 1.8. $\square$

However, application is not commutative.

Lemma 3.4. *The formula $s \cdot t : A \to t \cdot s : A$ is not valid in $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset models.*

Proof. Since the evidence set of a sum-term $s + t$ is defined to be *a* subset of the intersection of $E(w, s)$ and $(E.w, t)$, it is possible that $E(w, s \cdot t) \neq E(w, t \cdot s)$ and hence it is possible that only one of them is a subset of $[A]$. $\square$

4. Hyperintensionality

As mentioned by Artemov and Fitting [9], hyperintensionality is a key aspect of justification logic in contrast to standard modal logic, where it is missed.In modal logic, if A is logically equivalent to B, then $\Box A$ implies $\Box B$. This is not the case in justification logic where $A \leftrightarrow B$ and $s : A$ do not imply $s : B$. So justification logic is able to distinguish between equivalent contents. This is of importance when you think about propositions like '$0 = 0$' and Fermats last theorem. Both have the same content in a possible world semantics, namely all possible worlds. However, if some proof s is a justification for '$0 = 0$', s does not have to be a justification for Fermats last theorem as well.

In the standard semantics of justification logic, where the interpretation of a justification term is a sequence or set of formulas, hyperintensionality comes for free.

In subset semantics, however, *non-normal* worlds are required to obtain hyperintensionality. This can be seen easily in case of axioms and tautologies, that are necessarily true in each *normal* world. As long as you do not have *non-normal* worlds at hand, any justification term that justifies some tautology or some axiom, will justify all of them. The reason for this is simple: suppose you have a subset model $\mathcal{M} = (W, W_0, V, E)$ where $W_0 = W$, i.e. there are no *non-normal* worlds, and axioms A and B. Since A and B are axioms, they are true in all *normal* worlds and thus $[A] = [B] = W_0 = W$. So, since $[A] = [B]$, each subset of $[A]$ will be a subset of $[B]$ too and therefore each justification s that justifies A will justify B as well. So in this situation, there is no way to model a constant specification where different justifications actually justify different axioms without justifying all of them as a side effect.

By adding *non-normal* worlds we get hyperintensionality back. In *non-normal* worlds, axioms and tautologies do not have to evaluate to true, they might do so in some worlds but not in others. In a *non-normal* world two equivalent propositions may be evaluated to different truth values, and hence a justification that contains such a world may support one proposition but not the other. The importance of *non-normal* worlds for modelling hyperintensionality is worked out in detail by Jago in [20].

5. About D

In Chapters 1 and 2 we have presented families of justification logics and the corresponding subset models, which have been proven to be sound and complete. However, if **jd** was an axiom of the logic, completeness was only shown for axiomatically appropriate constant specifications. In this Chapter we will present slightly different subset models than in Chapter 2 which are complete for all constant specification even in presence of **jd**. The results of this Chapter are published in [18].

Before we start with the technical part, let us take a closer look at the D-axiom.[3] In justification logics there are several versions of axiom **jd**, based on various versions of the corresponding axiom in modal logic. In normal modal logics the following axioms are equivalent:

(1) $\neg\Box\perp$,

(2) $\Box A \to \neg\Box\neg A$,

(3) $\neg(\Box A \wedge \Box\neg A)$.

and can be translated into: there cannot be necessity for inconsistency. So, in modal logic which version you take is irrelevant.

However, as soon as we consider deontic logic, it really matters which one we take. By replacing $\Box$ with $\mathcal{O}$ and reading $\mathcal{O}A$ as "There is an obligation for A" the differences become visible. Then (1) should be read as "There can't be an obligation for an impossible state of affairs", however, (3) has another meaning: "There can't be an obligation for A and an obligation for $\neg A$". The difference is, that $(\mathcal{O}\perp)$ is conceptually impossible whereas $(\mathcal{O}A \wedge \mathcal{O}\neg A)$ is in principal logically possible, if there are two different obligations in the game: one for A and one for $\neg A$. Our daily life is full of such inconsistent obligations and philosophers have constructed even more such conflicts of duty. Worse still, these conflicts may even be derived

[3] The reflections in this Chapter are based on the work of Federico L.G. Faroldi and a remark of Meghdad Ghari.

from one and the same ethical principle, like for example the categorical imperative of Kant. So by following this principle we may be forced to do A and not to do A, like when we know where someone is hiding to avoid being murdered and the potential murderer asks us whether this person is at that place. Either we lie, which is strictly forbidden in Kant's interpretation of the categorical imperative, or we abet a murder, which also contradicts the categorical imperative. To sum up, in deontic logic it is of importance which version of D-axiom is chosen.

In justification logic, however, we distinguish between the terms. Having $(s : A \wedge t : \neg A)$ is not contradictory, as long as $s \neq t$. So in justification logic we can have an axiom $\neg(s : A \wedge s : \neg A)$ and still be able to deal with conflicts of duty.

The realization of (1) is the well known **jd**-axiom $\neg(t :\perp)$ and the realisation of (3) is known as **noc** : $\neg(t : A \wedge t : \neg A)$. We call the corresponding logical systems JD and JNoC. Corollary 5.12 establishes that the former implies the latter. The converse direction only holds in case of an axiomatically appropriate constant specification and the presence of $\mathbf{j+}$ which will be shown in Lemma 5.15, Lemma 5.16 and Remark 5.17. So there are two options to avoid the collapse between JD and JNoC : either skip $\mathbf{j+}$ or having a constant specification which is not axiomatically appropriate.

5.1. Syntax

We reuse the language $\mathcal{L}_J^{\mathsf{A}}$ as presented in Chapter 2.

The axioms of JD are the following:

cl	all axioms of classical propositional logic;
j	$s : (A \to B) \to (t : A \to s \cdot t : B)$;
j+	$s : A \vee t : A \to (s + t) : A$;
jd	$t :\perp \to \perp$.

The constant specification CS is defined analogue to Chapter 2 and we have the same rules (**MP**) and (**AN!**) as presented there.

5.2. D-arbitrary subset models

We present a novel class of subset models for JD and establish soundness and completeness. We refer to Chapter 2 and state that JD is part of the justification logics presented there. Hence by $\mathsf{JD}_{\mathsf{CS}}^{A}$-subset model we denote an $\mathsf{L}_{\mathsf{CS}}^{\mathsf{A}}$-subset model where L is JD.

Definition 5.1 (D-arbitrary subset model)**.** *D-arbitrary* CS*-subset models* $\mathcal{M} = (W, W_0, V, E)$ are defined like $\mathsf{JD}_{\mathsf{CS}}^{A}$-subset models with the condition in the definition of E:

$$\exists v \in W_0 \text{ with } v \in E(\omega, t)$$

being replaced with:

$$\exists v \in W_{\not\perp} \text{ with } v \in E(\omega, t)$$

where $W_{\not\perp} := \{\omega \in W \mid V(\omega, \perp) = 0\}$.

Definition 5.2 (D-arbitrary validity)**.** Let CS be a constant specification. We say that a formula F is *D-arbitrary* CS*-valid* if for each D-arbitrary CS-subset model $\mathcal{M} = (W, W_0, V, E)$ and each $\omega \in W_0$, we have $\mathcal{M}, \omega \Vdash F$.

We have soundness and completeness with respect to arbitrary constant specifications.

Theorem 5.3 (Soundness and Completeness). *Let* CS *be an arbitrary constant specification. For each formula* F *we have*

$$\mathsf{JD}_{\mathsf{CS}} \vdash F \quad \textit{iff} \quad F \textit{ is D-arbitrary } \mathsf{CS}\textit{-valid.}$$

The completeness proof is by a canonical model construction as in the case of $\mathsf{L}_{\mathsf{CS}}^{\mathsf{A}}$-subset models as presented in Chapter 2. We will only sketch main steps here. The canonical model is given as follows.

Definition 5.4 (Canonical Model)**.** Let CS be an arbitrary constant specification. We define the canonical model $\mathcal{M}^C = (W^C, W_0^C, V^C, E^C)$ by:

- $W^C = \mathcal{P}(\mathcal{L}_J)$.
- $W_0^C = \left\{ \Gamma \in W^C \;\middle|\; \Gamma \text{ is maximal } \mathsf{JD}_{\mathsf{CS}}\text{-consistent set of formulas} \right\}$.

- $V^C(\Gamma, F) = 1 \quad \text{iff} \quad F \in \Gamma$;
- $E^C(\Gamma, t) = \{\, \Delta \in W^C \mid \Delta \supseteq \Gamma/t \,\}$ where

$$\Gamma/t := \{F \in \mathcal{L}_J \mid t : F \in \Gamma\}.$$

The essential part of the completeness proof is to show that the canonical model is a D-arbitrary CS-subset model.

Lemma 5.5. *Let* CS *be an arbitrary constant specification. The canonical model* $\mathcal{M}^C$ *is a D-arbitrary* CS*-subset model.*

Proof. Let us only show the condition

$$\exists v \in W^C_{\not\perp} \text{ with } v \in E(\omega, t) \tag{5.1}$$

for all $\omega \in W_0$ and all terms t.

So let t be an arbitrary term and $\Gamma \in W^C_0$. Since Γ is a is maximal $\mathsf{JD}_{\mathsf{CS}}$-consistent set of formulas, we find $\neg(t : \perp) \in \Gamma$ and thus $t : \perp \notin \Gamma$. Let $\Delta := \Gamma/t$. We find that $\perp \notin \Delta$ and by definition $V^C(\Delta, \perp) = 0$. Thus $\Delta \in W^C_{\not\perp}$. Moreover, again by definition, $\Delta \in E^C(\Gamma, t)$. Thus (5.1) is established. ☐

Now the Truth lemma and the completeness theorem follow easily analogue to Lemma 2.9 and 2.10.

Remark 5.6. Our completeness result also holds in the setting with $\mathsf{c}^\star$. However, the proof that the canonical model is well-defined is a bit more complicated as one has to consider the case of $\mathsf{c}^\star$ separately.

5.3. No conflicts

So far, we have considered the explicit version of $\neg\mathcal{O}\perp$. In normal modal logic, this is provably equivalent to $\neg(\mathcal{O}A \wedge \mathcal{O}\neg A)$. In this section we study the explicit version of this principle, which we call NoC (*No Conflicts*), saying that reasons are self-consistent. That is A and $\neg A$ cannot be obligatory for one and the same reason. The axioms of JNoC are the axioms of JD where **jd** is replaced with:

noc $\quad \neg(t : A \wedge t : \neg A)$.

Accordingly, a constant specification for JNoC is defined like a constant specification for JD except that the constants justify axioms of JNoC.

Given a constant specification CS for JNoC, the logic $\mathsf{JNoC}_{\mathsf{CS}}$ is defined by the axioms of JNoC and the rules modus ponens and axiom necessitation.

Definition 5.7 (NoC subset model)**.** A *NoC* CS*-subset model*

$$\mathcal{M} = (W, W_0, V, E)$$

is defined like a general subset model with the condition in the definition of E:

$$\exists v \in W_0 \text{ with } v \in E(\omega, t)$$

being replaced with:

$$\exists v \in W_{\mathsf{nc}} \text{ with } v \in E(\omega, t)$$

where

$$W_{\mathsf{nc}} := \{\omega \in W \mid \text{for all formulas A } (V(\omega, A) = 0 \text{ or } V(\omega, \neg A) = 0)\}.$$

The notion of NoC CS-validity is now as expected.

Definition 5.8 (NoC validity)**.** Let CS be a constant specification. We say that a formula F is *NoC* CS*-valid* if for each NoC CS-subset model $\mathcal{M} = (W, W_0, V, E)$ and each $\omega \in W_0$, we have $\mathcal{M}, \omega \Vdash F$.

Theorem 5.9 (Soundness and Completeness). *Let* CS *be an arbitrary constant specification. For each formula* F *we have*

$$\mathsf{JNoC}_{\mathsf{CS}} \vdash F \quad \textit{iff} \quad F \textit{ is NoC } \mathsf{CS}\textit{-valid.}$$

Again the completeness proof uses the canonical model construction from Definition 5.4 except that we set

$$W_0^C = \{\Gamma \in W^C | \Gamma \text{ is maximal } \mathsf{JNoC}_{\mathsf{CS}}\text{-consistent set of formulas}\}.$$

Now we have to show that the defined structure is an NoC CS-subset model.

Lemma 5.10. *Let* CS *be an arbitrary constant specification. The canonical model* $\mathcal{M}^C$ *is an NoC* CS*-subset model.*

Proof. As before, we only show the condition

$$\exists \upsilon \in W^C_{\mathsf{nc}} \text{ with } \upsilon \in E(\omega, t) \tag{5.2}$$

for all $\omega \in W_0$ and all terms t.

So let t be an arbitrary term and $\Gamma \in W^C_0$. Let A be an arbitrary formula. Since Γ is a is maximal $\mathsf{JNoC}_{\mathsf{CS}}$-consistent set of formulas, we find

$$\neg(t : A \wedge t : \neg A) \in \Gamma$$

and thus $t : A \wedge t : \neg A \notin \Gamma$. Thus, again by maximal consistency,

$$t : A \notin \Gamma \text{ or } t : \neg A \notin \Gamma.$$

Let $\Delta := \Gamma/t$. We find that

$$A \notin \Delta \text{ or } \neg A \notin \Delta$$

and hence, by definition,

$$V^C(\Delta, A) = 0 \text{ or } V^C(\Delta, \neg A) = 0.$$

Thus $\Delta \in W^C_{\mathsf{nc}}$. Moreover, again by definition, $\Delta \in E^C(\Gamma, t)$. Therefore (5.2) is established: $\square$

Again the Truth lemma and the completeness theorem follow easily.

5.4. Formal comparison

The following lemmas establish the exact relationship between JD and JNoC. First we show that $\mathsf{JD}_{\mathsf{CS}}$ proves that reasons are consistent among them, i.e. that $\neg(s : A \wedge t : \neg A)$ holds for arbitrary terms s and t, which is the consistency principle used in [17].

Lemma 5.11. *Let* CS *be an arbitrary constant specification. Then* $\mathsf{JD}_{\mathsf{CS}}$ *proves* $\neg(s : A \wedge t : \neg A)$ *for all terms* s, t *and all formulas* A.

Proof. Suppose towards a contradiction that $s : A \wedge t : \neg A$. Thus we have $s : A$ and $t : \neg A$ where the latter is an abbreviation for $t : (A \to \bot)$ (by the definition of the symbol $\neg$). Thus using axiom **j**, we get $t \cdot s : \bot$ and by axiom **jd** we conclude $\bot$. $\Box$

Corollary 5.12. *For any constant specification* CS*,* $\mathsf{JD}_{\mathsf{CS}}$ *proves every instance of* ***noc****.*

Remark 5.13. It is only by coincidence that Lemma 5.11, and thus also Corollary 5.12, hold for arbitrary constant specifications. If we base our propositional language on different connectives (say $\wedge$ and $\neg$ instead of $\rightarrow$ and $\bot$), then Lemma 5.11 and Corollary 5.12 only hold for axiomatically appropriate constant specifications.

The proof of Lemma 5.11 then is as follows. Since CS is axiomatically appropriate, there exists a term r such that

$$r : (\neg A \rightarrow (A \rightarrow \bot)) \tag{5.3}$$

is provable where $\bot$ is defined as $P \wedge \neg P$ (for some fixed P) and $F \rightarrow G$ is defined as $\neg(F \wedge \neg G)$. From (5.3) and axiom **j** we get

$$t : \neg A \rightarrow r \cdot t : (A \rightarrow \bot).$$

Thus from $s : A \wedge t : \neg A$, we obtain $(r \cdot t) \cdot s : \bot$, which contradicts axiom **jd** as before.

Next we show that also $\mathsf{JNoC}_{\mathsf{CS}}$ proves that reasons are consistent among them.

Lemma 5.14. *Let* CS *be an arbitrary constant specification. Then* $\mathsf{JNoC}_{\mathsf{CS}}$ *proves* $\neg(s : A \wedge t : \neg A)$ *for all terms* s, t *and all formulas* A*.*

Proof. Suppose towards a contradiction that $s : A \wedge t : \neg A$ holds. Using axiom **j+** we immediately obtain $s + t : A \wedge s + t : \neg A$. By axiom **noc** we conclude $\bot$, which establishes $\neg(s : A \wedge t : \neg A)$. $\Box$

Next we show that $\mathsf{JNoC}_{\mathsf{CS}}$ with an axiomatically appropriate constant specification proves $\neg(t : \bot)$.

Lemma 5.15. *Let* CS *be an axiomatically appropriate constant specification. Then* $\mathsf{JNoC}_{\mathsf{CS}}$ *proves* $\neg(t : \bot)$ *for each term* t*.*

Proof. Because CS is axiomatically appropriate, there are terms r and s such that
$$r : (\bot \to P) \quad \text{and} \quad s : (\bot \to \neg P).$$
Therefore, we get
$$t : \bot \to r \cdot t : P \quad \text{and} \quad t : \bot \to s \cdot t : \neg P.$$
Thus we have $t : \bot \to (r \cdot t : P \wedge s \cdot t : \neg P)$. Together with the previous lemma, this yields $t : \bot \to \bot$, which is $\neg(t : \bot)$. □

Here the requirement of an axiomatically appropriate constant specification is necessary.

Lemma 5.16. *There exists a NoC* CS*-subset model* $\mathcal{M} = (W, W_0, V, E)$ *with some* $\omega \in W_0$ *such that*
$$\mathcal{M}, \omega \Vdash t : \bot$$
for some term t*.*

Proof. Consider the empty CS and the following model:

(1) $W = \{\omega, v\}$ and $W_0 = \{\omega\}$

(2) $V(v, \bot) = 1$ and $V(v, F) = 0$ for all other formulas F

(3) $E(\omega, t) = \{v\}$ for all terms t.

We observe that $v \in W_{\mathsf{nc}}$. So the model is well-defined. Further, we find $E(\omega, t) \subseteq [\bot]$. Since $\omega \in W_0$, we get $V(\omega, t : \bot) = 1$. We conclude
$$\mathcal{M}, \omega \Vdash t : \bot.$$
□

Remark 5.17. For Lemmas 5.14 and 5.15, the presence of the $+$ operation is essential. Consider a term language without $+$ and the logic JNoC^- being JNoC without $\mathsf{j+}$. Let CS be an axiomatically appropriate CS for JNoC^-. There is a NoC CS-subset model $\mathcal{M}$ for $\mathsf{JNoC}^-_{\mathsf{CS}}$ with a normal world ω such that
$$\mathcal{M}, \omega \Vdash s : P \wedge t : \neg P$$
for some terms s and t and some proposition P.

Hence, if we drop the + operation, we can have self-consistent reasons without getting reasons that are consistent among them even in the presence of an axiomatically appropriate constant specification.

Instead of using an axiomatically appropriate constant specification, we could also add the schema $s : \top$ to $\mathsf{JNoC}_{\mathsf{CS}}$ in order to derive **jd**.

Lemma 5.18. *Let* CS *be an arbitrary constant specification. Let* $\mathsf{JNoC}^+_{\mathsf{CS}}$ *be* $\mathsf{JNoC}_{\mathsf{CS}}$ *extended by the schema* $s : \top$ *for all terms* s*. We find that*

$$\mathsf{JNoC}^+_{\mathsf{CS}} \vdash \neg(t : \bot) \quad \textit{for each term } t.$$

Proof. The following is an instance of axiom **noc**

$$\neg(t : \bot \wedge t : \neg\bot).$$

Using the definition $\top := \neg\bot$ and propositional reasoning, we obtain

$$t : \top \to \neg(t : \bot).$$

Using $t : \top$ and modus ponens, we conclude $\neg(t : \bot)$. $\square$

Part II.

New Operators and Terms for Justification Logics

In Part I we have shown how subset models can be used to model standard justification logics. In this part, we show how they can model new operators and terms and how they can be used to study dynamic processes.

In Chapter 6 we show how Artemov's approach to aggregated evidence can be subsumed by subset models, i.e. how we can find the best lower bound for the probability that X if we have probabilities for $F_1, \dots F_n$ and $F_1, \dots F_n \vdash X$.

In Chapter 7 we explore various kinds of combining justifications. We analyse how two (or more) justifications can interact together and introduce some new operators to model these interactions.

In Chapter 8 we introduce new types of terms that include primary beliefs. We use them to model things like intuition, where we have beliefs for which we do not have any explicit primary cognitive processes, and selective perceptions, i.e. perceptions where we ignore those parts, which contradict our presumptions.

In Chapter 9 we investigate a version of contraction and how subset models can be adapted for it.

Finally, in Chapter 10 we analyse dynamic aspects of subset models and introduce the new term $\mathsf{up}(A)$. This term identifies the updating process as a specific justification and we will read the formula $[A]\mathsf{up}(A) : A$ as *after an update with A, this update is a justification to believe A.* We study the main properties of the resulting logic as well as the differences to a previous (symbolic) approach to belief expansion in justification logic.

6. Artemov's aggregated evidence and $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset models

It was our goal to use subset models for justification logics with probabilities. The most promising strategy was to introduce a probability measure on evidence sets. This is the same approach that Artemov took in [7]. We therefore used his work as a guideline to adapt our subset models in an appropriate way.

6.1. Aggregated evidence

Artemov considers the case in which we have a database, i.e. a set of propositions $\Gamma = \{F_1, \ldots F_n\}$ with some kind of probability estimates and in which we also have some proposition X that logically follows from Γ. Then we can search for the best justified lower bound for the probability of X. He presents us with a nice way to find this lower bound. To find it, he introduces probability events $u_1, \ldots, u_n$, each of which supports some proposition in Γ, i.e. $u_i : F_i$, and calculates some aggregated evidence $e(u_1, \ldots, u_n)$ for X with them. The probability of e then provides a tight lower bound for the probability of X.

The trick he uses is the following:

(1) First he collects all subsets Δ_i of Γ which support X, i.e. $\Delta_i \vdash X$, and creates a new evidence t_i from all the corresponding u_{i_j} s.t. $u_{i_j} : F_{i_j}$ for each $F_{i_j} \in \Delta_i$.

(2) In a second step he combines all these new pieces of evidence to a new piece of evidence (the so-called aggregated evidence) that actually is the greatest evidence supporting X.

The model he has in mind contains some evaluation in a probability space $(\Omega, \mathcal{F}, P)$ with a mapping $\star$ from propositions to Ω and evidence terms to $\mathcal{F}$ that meets some restrictions (for more details on this see [7]). Step (1) is to create a new piece of evidence t_i for each Δ_i described above, which consists of the intersection of the corresponding u_{i_j}'s.

$$t_i := \bigcap\{u_{i_j} \mid u_{i_j} \subseteq F^{\star}_{i_j} \text{ for some } F_{i_j} \in \Delta_i\}.$$

Step (2) then is to unite all these pieces of evidence to a new so-called aggregated evidence:

$$\mathsf{AE}^{\Gamma}(\mathsf{X}) := \bigcup\{t_i \mid t_i \text{ is an evidence for } X \text{ obtained by step (1)}\}.$$

On the syntactic side evidence terms are built from variables $u_1, \ldots, u_m$, constants 0 and 1 and operations $\cap$ and $\cup$, where st is used as an abbreviation for $s \cap t$. With this we can built a free distributive lattice $\mathcal{L}_n$ where st is the meet and $s \cup t$ is the join of s and t, 0 is the bottom and 1 the top element of this lattice. Moreover Artemov defines formulas in a usual way from propositional letters $p, q, r, \ldots$ by the usual connectives and adds formulas of the kind $t : F$ where t is an evidence term and F a purely propositional formula.

The logical postulates of the logic of Probabilistic Evidence PE are:

A1 axioms and rules of classical logic in the language of PE;

A2 $s : (A \to B) \to (t : A \to [st] : B)$;

A3 $(s : A \wedge t : A) \to [s \cup t] : A$;

A4 $1 : A$, where A is a propositional tautology,
$0 : F$, where F is a propositional formula;

A5 $t : X \to s : X$, for any evidence terms s and t such that $s \preceq t$ in $\mathcal{L}_n$.

Artemov presents Soundness and Completeness proofs connecting PE with the presented semantic, for more details see [7].

6.2. Subset models for PE

Before we can start adapting Artemov's approach to our models, we have to point out some differences between the semantics and syntax used. First, contrary to the models of Artemov, subset models may contain non-normal worlds, but this does not significantly affect the applicability of Artemov's approach on them.

Another difference is that our evidence function has a different domain. In Artemov's models the evidence functions is $E : \mathsf{Tm} \to \mathcal{P}(\Omega)$ while in our models it is $E : W \times \mathsf{Tm} \to \mathcal{P}(W)$. This difference is due to the fact that we allow terms to justify non-purely propositional formulas. Although we need to adapt Artemov's definitions, these adaptations will maintain the essential characteristics. So let us adapt the $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset models to aggregated $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset models by first describing the new syntax for the terms:

Definition 6.1 (Justification Terms)**.** Justification terms are built from constants $0, 1, c_i$ and variables x_i and the special and unique constant $\mathsf{c}^{\star}$ according to the following grammar:

$$t ::= 0 \mid 1 \mid c_i \mid x_i \mid \mathsf{c}^{\star} \mid (t + t) \mid (t \cup t) \mid !t$$

This set of terms is denoted by Tm^{P}. As before, we introduce the abbreviation $st := s + t + \mathsf{c}^{\star}$.

Even though we have other operators as well, we can construct a free distributive lattice where we take $s + t$ as the meet of s and t, $s \cup t$ as the join of them, 0 as the bottom element of the lattice. Note that st then is the meet of s, t, and $\mathsf{c}^{\star}$. Moreover, 1 and $!t$ are treated like constants.[4] As usual, we have

$$s \preceq t \quad \text{iff} \quad s \cup t = t \tag{6.1}$$

Consequently not all pairs of terms are comparable. This, however, does not have any implications so far.

There is no difference to our subset models regarding the rules for forming formulas except that the terms are contained in Tm^{P}, of course. The set

[4] We do not claim that 1 is the top element since some set $E(\omega, t)$ for a world $\omega \in W_0$ and $t \in \mathsf{Tm}^{\mathsf{P}}$ may contain non-normal worlds. If we claimed that 1 was the top element we would obtain $t \preceq 1$ and furthermore the set $E(\omega, 1)$ would contain non-normal worlds as well. But since in non-normal worlds axioms may not be true, $E(\omega, 1) \not\subseteq [A]$ for some axiom A may be the case and therefore axiom **A4** would fail.

of formulas built according to this grammar and these rules is denoted by $\mathcal{L}_{prob}$.

In the definition of $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset models we only change the conditions on the evidence function and the domain of V.

Definition 6.2 (PE-adapted subset models)**.** An $\mathsf{L}^{\star}_{\mathsf{CS}}$- subset model is called a PE-adapted $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset model if the valuation function and the evidence function meet the additional conditions respectively are redefined as follows:

- $V : W \times \mathcal{L}_{prob} \to \{0, 1\}$ where all conditions listed in Definition 1.2 remain the same.
- For all $\omega \in W_0$ and for all $s, t \in \mathsf{Tm}^{\mathsf{P}}$:
  - $E(\omega, 1) = W_0$;
  - $E(\omega, 0) = \emptyset$;
  - $E(\omega, s \cup t) = E(\omega, s) \cup E(\omega, t)$.

In fact, such an PE-adapted $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset model is a model of probabilistic evidence PE.

Definition 6.3 (PE-validity)**.** Let CS be a constant specification. We say that a formula F is PE-*valid* if for each PE-adapted $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset model $\mathcal{M} = (W, W_0, V, E)$ and each $w \in W_0$, we have $\mathcal{M}, w \Vdash F$.

Theorem 6.4 (Soundness). *For any formula $F \in \mathcal{L}_{prob}$ we have that if* $\mathsf{PE} \vdash F$*, then F is* PE*-valid.*

Proof. The proof is by induction on the length of the derivation of F:

- If F is derived by axiom necessitation or modus ponens or is an instance of axiom **A1**, then the proof is the analogue as in Theorem 1.6 since the relevant definitions have remained the same.
- If F is an instance of axiom **A2**, the proof is analogue to the proof of Lemma 1.8: Suppose $\mathcal{M}, \omega \Vdash s : (A \to B)$ and $\mathcal{M}, \omega \Vdash t : A$ then $E(\omega, s) \subseteq [A \to B]$ and $E(\omega, t) \subseteq [A]$.

$$E(\omega, st) = E(\omega, s + t + \mathsf{c}^{\star}) \subseteq$$
$$E(\omega, s) \cap E(\omega, t) \cap E(\omega, \mathsf{c}^{\star}) \subseteq [A \to B] \cap [A] \cap E(\omega, \mathsf{c}^{\star}).$$

Hence, for all $\upsilon \in E(\omega, st)$ we have $V(\upsilon, A \to B) = 1$ and $V(\upsilon, A) = 1$ and $\upsilon \in E(\omega, \mathsf{c}^\star)$ and therefore $V(\upsilon, B) = 1$. Hence $E(\omega, st) \subseteq [B]$ and we obtain $\mathcal{M}, \omega \Vdash st : B$.

- If F is an instance of axiom **A3**, then $F = (s : A \wedge t : A) \to [s \cup t : A]$ for some $A \in \mathcal{L}_{prob}$, $s, t \in \mathsf{Tm}^{\mathsf{P}}$. Suppose $\mathcal{M}, \omega \Vdash s : A \wedge t : A$ hence $E(\omega, s) \subseteq [A]$ and $E(\omega, t) \subseteq [A]$. Therefore

$$E(\omega, s \cup t) \subseteq E(\omega, s) \cup E(\omega, t) \subseteq [A]$$

  and since $\omega \in W_0$ we obtain $\mathcal{M}, \omega \Vdash s \cup t : A$.

- If F is an instance of axiom **A4**, then either $F = 1 : A$ for some axiom A or $0 : G$ for some formula G.
  Suppose $F = 1 : A$ for some axiom A. We assume that $\mathcal{M}, \omega \Vdash A$ for all $\omega \in W_0$, hence $E(\omega, 1) = W_0 \subseteq [A]$ and therefore $\mathcal{M}, \omega \Vdash 1 : A$ for all $\omega \in W_0$.
  Suppose $F = 0 : G$: For any $\omega \in W_0$ we have $E(\omega, 0) = \emptyset$ by Definition 6.2. Since $\emptyset$ is a subset of any subset of W, we obtain $E(\omega, 0) = \emptyset \subseteq [G]$ for any formula $G \in \mathcal{L}_{prob}$.

- F is an instance of axiom **A5**. Assume $\mathcal{M}, \omega \Vdash t : X$ for some term t and some formula X and let $s \preceq t$. By (6.1) we find $t = s \cup t$. Thus

$$E(\omega, t) = E(\omega, s \cup t) = E(\omega, s) \cup E(\omega, t)$$

  and therefore $E(\omega, s) \subseteq E(\omega, t)$. The assumption $\mathcal{M}, \omega \Vdash t : X$ means that $E(\omega, t) \subseteq [X]$. Hence we also get $E(\omega, s) \subseteq [X]$ and conclude $\mathcal{M}, \omega \Vdash s : X$. □

Theorem 6.5 (model existence). *There exists a* PE*-adapted* $\mathsf{L}^\star_{\mathsf{CS}}$*-subset model.*

Proof. We construct a model $\mathcal{M} = (W, W_0, V, E)$ as follows:

- $W = W_0 = \{\omega\}$.
- The valuation function is built bottom up:
  (1) $V(\omega, \bot) = 0$;
  (2) $V(\omega, P) = 1$, for all $P \in \mathsf{Prop}$;

(3) $V(\omega, A \to B) = 1$ iff $V(\omega, A) = 0$ or $V(\omega, B) = 1$;

(4) $V(\omega, t : F) = 1$ iff $t \not\geq 1$ or if $t \geq 1$ and $V(\omega, F) = 1$.

- $E(\omega, t) = \begin{cases} \{\omega\} & \text{if } t \geq 1 \\ \emptyset & \text{otherwise.} \end{cases}$

It is straightforward to show that $\mathcal{M}$ is indeed a PE-adapted $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset model. Let us only show the condition $E(\omega, s \cup t) = E(\omega, s) \cup E(\omega, t)$.

Suppose first $s, t \not\geq 1$, Then $E(\omega, s \cup t) = \emptyset = E(\omega, s) = E(\omega, t)$ and hence the claim follows immediately.

Suppose at least one term of s and t is in greater than 1, then we obtain that $E(\omega, s) = \{\omega\}$ or $E(\omega, t) = \{\omega\}$ and hence $E(\omega, s) \cup E(\omega, t) = \{\omega\}$ and since $s \leq s \cup t$ and $t \leq s \cup t$ we obtain $s \cup t \geq 1$ and therefore $E(\omega, s \cup t) = \{\omega\}$, so the claim holds. $\square$

Note that we cannot use the canonical model to show that adapted subset models exists since in the canonical model

$$E(\Gamma, s \cup t) \not\subseteq E(\Gamma, s) \cup E(\Gamma, t).$$

However, in an adapted subset model we need these sets to be equal (see Definition 6.2) since otherwise axioms **A3** and **A5** would not be sound.

7. Non-monotone combinations of justifications

In $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset models the operator $+$ is monotone and so is the application-operator too, i.e.

$$s : A \to s + t : A$$

$$s : A \to s \cdot t : A.$$

This means that by combining several justifications, we never lose information. The difference between both versions of subset models with respect to this aspect of the application-operator was studied in Chapter 3.

If we want to model self consistent reasons without getting reasons that are as well consistent among them, we have to drop the **j+**-axiom, as we have seen in Remark 5.17. That seems to be intuitive, since **j+** models a way of combining justifications that have unbreakable validity, as it is the case for mathematical proofs. But if we think about less reliable justifications, which may be misleading, it should be possible to have justifications for A and others for $\neg A$. A standard example is having two volumes of an encyclopedia s and t. Let us assume that the first one, s, offers a justification for A and the other one, t, contains a correction to the first volume saying that $\neg A$. Hence although s offers a justification for A, we would not say that any combination $s + t$ is still a justification for A. Even in our daily lives we encounter many justifications which lead to contradictions among each other and one of our daily challenges is to learn how to deal with these contradictions. So if we want to model how we combine justifications and how we use them to form our beliefs, it seems to be essential to rethink the standard **j+**-axiom and its monotonicity.

When giving up monotonicity, we may even consider cases where

$$s : A \wedge t : A \to s + t : A$$

does not hold. This may seem a little strange at first. However, such pairs of incompatible justifications exist. Consider the following historical example. In the 19th century the feminist movement in Europe was divided into two fractions according to the justifications for the right of women to political participation: Egalitarism and Differentialism. The first fraction argued that since women are equal to men, they should be subject to the Human Rights as well. Hence if we have $s =$ "women are equal to men (and hence should have the same rights)" and $A =$ "women should have the right to vote" the egalitarians claimed $s : A$. The other fraction used the opposite justification to justify A, namely $t =$ "women are different to men (and hence their experience will be an enrichment)" and concluded $t : A$. So we have two justifications that justify the same conviction but which are not compatible and hence cannot be used together to justify anything.[5]

To model non-monotone combinations of justifications, we will introduce a plausibility relation $\succcurlyeq$ on the worlds to order them. That means we say that some of the worlds are more plausible than others. When we consider non-monotone justifications, it is clear that we do not mean justifications as proofs but rather justification as empirical observations which support some beliefs. And if we ask whether or not some observation t supports a belief A, we will only consider those worlds in t that are the most plausible ones. If they all support the belief that A, then we say $t : A$.

If one thinks about plausibility, then there are two main possibilities to model the term *most plausible*. One is that the set of the most plausible worlds contains the worlds that are more plausible than all others, i.e.

$$\mathrm{opt}_{\succcurlyeq}(S) = \{x \in S : x \succcurlyeq y \text{ for all } y \in S\}.$$

In this case all most plausible worlds are $\succcurlyeq$-comparable to each other. The other way is by demanding that for each most plausible world there does not exist a strictly more plausible world, i.e.

$$\max_{\succcurlyeq}(S) = \{x \in S : y \succcurlyeq x \to x \succcurlyeq y \text{ for all } y \in S\}.$$

[5] Even today these two main lines of argumentation exist within feminist movements. But nowadays the concept of gender and the reflection on its social construction process refined the argumentation a lot, so that combinations of variants of s and t are possible. For more details to the feminist movement in Europe see [27].

By using this definition for *most plausible*, two non-comparable worlds may be in the set of the most plausible worlds. For our purpose it seems to be more reasonable to consider $\max_{\succcurlyeq}$ rather than $\mathrm{opt}_{\succcurlyeq}$ because we see no necessity to demand comparability of all most plausible worlds. For more details on the differences of these two ways of defining most plausible worlds see Parent [29].

As an example of such a plausibility relation, consider a fan of Ockham's razor. This agent will always prefer explanations that need the fewest added entities. So within an observation that may serve as a justification for something, there may be a world w_0 that needs no metaphysics, another one w_1 that needs some added entities and a third one w_2 that needs a whole cabinet of extra ontological entities. Then this agent will have $w_0 \succcurlyeq w_1 \succcurlyeq w_2$. If this agent is lucky enough to observe a supernova in the Milky Way, she will take this observation as justification for the fact that a star has exploded and ignore the fact that this observation could also indicate the birth of a god.

7.1. New Operators

So far we only used standard operators on justification terms like $\cdot, +, \cup$. But there seem to be more options if we investigate the combination of justification based on observations. We will model them by two new operators: $\hat{+}$, $\grave{+}$ and refine the meaning of the already used ones. Hence we have these five ways how an agent can combine justification terms:

- $s \cdot t$: The meaning is the same as in $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-models, i.e. applying modus ponens on the formulas justified by s with the formulas justified by t where the order of these two sets of formulas matters. So, the agent takes two steps: first she considers all the formulas of kind $A \to B$ that are true in all most plausible worlds of observation s and compares them with all the formulas of kind A in the most plausible worlds of t. Then she focuses only on worlds in which the corresponding B is true. What is true in the most plausible worlds of $[B]$ will be justified by $s \cdot t$. This is a non-commutative way of combining observations. If the agent wants to say that she considers all formulas that are justified by applying modus ponens on all the formulas in s and t she will have a justification term $(s \cdot t) \circ (t \cdot s) \circ (s \cdot s) \circ (t \cdot t)$ where $\circ$ can be any of $+$, $\hat{+}$, $\cup$.

- $s+t$ and $s\hat{+}t$: In both cases the agent will consider facts that are true in the most plausible worlds that are part of both justification. The difference is whether she is willing to accept that the combination of s and t justifies facts that are true only if we make compromises in the plausibility of the worlds in which these facts are true. Suppose for simplicity that s and t both leave us with only three possible worlds: $s = \{w_1, w_2\}$ and $t = \{w_1, w_3\}$ and we have that A is true in all worlds of s and B is true in all worlds of t. Furthermore, $w_2 \succcurlyeq w_1$ and $w_3 \succcurlyeq w_1$. Now we can use intersection of s and t in two different ways: either we accept all facts that are true in the common worlds, i.e. w_1, and therefore we say that the combination of s and t justifies both, i.e. $s + t : A$ and $s + t : B$, or we say that the combination of s and t only justifies the facts that are true in the intersection of the most plausible worlds of s and t. In the latter case this combination justifies nothing. So it is all about the agent's willingness to make compromises in the level of plausibility.

  Let us consider again the fan of Ockham's razor. If she makes two observations s and t which together suggest that additional entities are needed (A="there is a God"), will she be willing to add these entities to her beliefs or will she say that at least one of the two observations cannot be correct? We have here that $s + t : A$ is true but $s \hat{+} t : A$ is false.

  In the example of the feminist movement we have two justifications that we cannot combine so that the combination still supports anything that is plausible. So there as well we have $s + t : A$ but not $s \hat{+} t : A$ since there is no common most plausible world in s and t.

- $s \cup t$: We say that $s \cup t$ justifies A only if A is true in the most plausible worlds of the union of both justifications. Please keep in mind that this does not mean that necessarily $s : A$ or $t : A$. To see this consider the example from Figure 1. There is a world w_1, which is maximal in s and where A is not true but there also is a world w_2 in t in which A is true and which is more plausible than w_1, and likewise worlds w_4 and w_5.

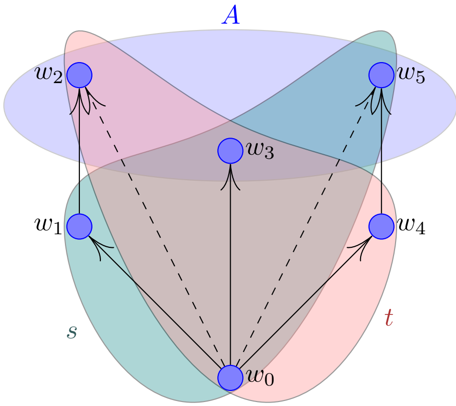

Figure 1.: Example, where the most plausible worlds of the union of s and t are A-worlds but where neither all most plausible worlds of s nor all most plausible worlds of t are A-worlds.

Let us consider, for example, the role of a member of a jury at a court hearing. This member, let us name her Anne, hears different testimonies to decide A="the accused is guilty". She may not be able to classify all the witnesses as credible, but let's assume that there are two testimonies s and t that she herself considers credible. Let us now look at the first testimony of witness s. If this testimony leaves open doubts about the guilt of the defendant, then these doubts should lead to Anne's recognition of innocence in the sense of *in dubio pro reo*, unless the doubt is removed by the second testimony t, by giving additional clues or by providing better explanations for an aspect that seem to be relieving in s. The same applies, of course, to t with regard to s. Anne will not necessarily decide whether s or t is right or if one is right at all. But she will take any remaining doubts of her credible witnesses as an objection to the verdict guilty. If, however, the two testimonies s and t united in this sense leave no doubt about guilt, $s \cup t : A$ applies.

- $s \grave{+} t$: This is a non-commutative but monotone way of combining two justifications. It is to have a justification s on which one relies so much that one only accepts other justification t if they have common most plausible worlds and we ignore t, if not.[6]

  If my best friend, who has never lied, who is very smart, who knows about investments and who is a very conscientious person, tells me that A is a good investment, I will believe him, even if others say the opposite. All situations in which people are resistant to consulting are of this type.

7.2. Syntax

To model combined justifications of this kind, we have to adapt our framework. We will work with our second kind of subset models: $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-subset models instead of $\mathsf{c}^{\star}$-subset models since in the latter the application operator only works with a monotone sum-operator.[7]

Justification terms are built from constants c_i and variables x_i according to the following grammar:

$$t ::= c_i \mid x_i \mid (t \cdot t) \mid (t + t) \mid (t \mathbin{\hat{+}} t) \mid (t \cup t) \mid (t \mathbin{\grave{+}} t)$$

This set of terms is denoted by $\mathsf{Tm}^{\mathsf{nm}}$. The operations $\cdot, +, \hat{+}, \cup$ and $\grave{+}$ are left-associative.

Formulas are built from atomic propositions p_i and the following grammar:

$$F ::= p_i \mid \perp \mid F \to F \mid t : F$$

The set of atomic propositions is denoted by Prop and the set of all formulas is denoted by $\mathcal{L}_J^{nm}$. Again we use the other logical connectives as abbreviations.

The logic L^{nm} is as follows: As usual we have two sets of axioms. The

[6] Of course we could do the similar thing for t and hence define a justification $s \acute{+} t$. But this would not really add something new.

[7] With such a monotone sum-operator we have that $s : A \to s + t + \mathsf{c}^{\star} : A$ and $t : (A \to B) \to s + t + \mathsf{c}^{\star} : (A \to B)$ so that $\mathsf{c}^{\star}$-part can apply modus ponens to derive $s + t + \mathsf{c}^{\star} : B$.

first one is $\mathsf{L}^{\mathsf{nm}}_{\alpha}$ with $s, t \in \mathsf{Tm}^{\mathsf{nm}}$ and $A, B, C \in \mathcal{L}^{nm}_{J}$:

cl	all axioms of classical propositional logic;	
j	$s : (A \to B) \to (t : A \to s \cdot t : B)$;	
NM1	$s \cup t : A \wedge s \hat{+} t : C \to s \hat{+} t : A$	for a $C \in \mathcal{L}^{nm}_{J}$
NM2	$s + t : A \wedge s \hat{+} t : C \to s \hat{+} t : A$	for a $C \in \mathcal{L}^{nm}_{J}$
NM3	$s : A \wedge t : A \to s \cup t : A$	
NM4	$(s : A \vee t : A) \wedge s \hat{+} t : C \to s \hat{+} t : A$	for a $C \in \mathcal{L}^{nm}_{J}$
NM5	$s : A \to s \grave{+} t : A$	
NM6	$s \grave{+} t : A \wedge s \hat{+} t : C \to t \grave{+} s : A$	for a $C \in \mathcal{L}^{nm}_{J}$

The other is denoted by $\mathsf{L}^{\mathsf{nm}}_{\beta}$ and contains:

jd	$t : \bot \to \bot$;
jt	$t : A \to A$.

By L^{nm} we denote all logics that are composed from the whole set of $\mathsf{L}^{\mathsf{nm}}_{\alpha}$ and some subset of $\mathsf{L}^{\mathsf{nm}}_{\beta}$.

As in Chapter 1 we assume having a constant specification CS:

$$\mathsf{CS} \subseteq \{(c, A) \mid c \text{ is a constant and } A \text{ is an axiom of } \mathsf{L}^{\mathsf{nm}}\}$$

In this sense ${\mathsf{L}^{\mathsf{nm}}}_{CS}$ denotes the logic L^{nm} with the constant specification CS. To deduce formulas in ${\mathsf{L}^{\mathsf{nm}}}_{CS}$ we use again a Hilbert system given by L^{nm} and the rules modus ponens and and axiom necessitation

$$\frac{}{c : A}\ (\mathrm{AN}) \quad \text{where } (c, A) \in \mathsf{CS}$$

NM1 says that if A is supported by all most possible worlds of the union of two justification and these two justification have an non-empty set of common most plausible worlds $s \hat{+} t$ this set supports as well that A. Or to come back on the tribunal example: if two testimonies are compatible and if, taken together by considering all observations they entail, they are a justification for A, they are as well a justification for A if we only consider those observations on which both witnesses agree.

NM2 says that facts which are supported by the combination of two compatible justifications, provided one is willing to reduce the plausibility level, are supported even without this reduction.

NM3 says that if two justification s and t each justify a fact A and hence A is true in all most plausible worlds of both justifications, it is also true in all most plausible worlds that are in the union. So we refer again on our tribunal example: both testimonies justify that the accused is guilty, hence taking all parts of their testimonies together still justifies that the accused is guilty.

NM4 says that if a fact is true in all most plausible worlds of a justification s and s is compatible with some other justification t such that one does not have to make a compromise in plausibility, then these two justification together still justify A without compromise on plausibility.

NM5 just says, that non-defeatable justification never lose their ability to justify what they justify. Of course, the arrow is only from left to right because if t is compatible with s it may refine the worlds we consider and hence widen the set of facts that are supported by the combination.

NM6 says, that if two justifications s and t are compatible, they together support the same facts regardless which of them is seen as more reliable.

7.3. Semantics

Definition 7.1 (Non-monotone subset models)**.** Given some L^{nm}-logic and some constant specification CS, then a corresponding $\mathsf{L}^{\mathsf{nm}}{}_{\mathsf{CS}}$-non-monotone subset model $\mathcal{M} = (W, W_0, \succcurlyeq, V, E)$ is defined as follows:

- W, W_0 are defined like in $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-subset models.
- $\succcurlyeq$ is a preorder on W, so:
  - $w \succcurlyeq w$ for all $w \in W$ (reflexivity);
  - if $w \succcurlyeq v$ and $v \succcurlyeq u$ then $w \succcurlyeq u$ (transitivity).
- $V : W \times \mathcal{L}_J^{nm} \to \{0,1\}$ s.t for all $\omega \in W_0$, $t \in \mathsf{Tm}^{\mathsf{nm}}$, $F, G \in \mathcal{L}_J^{nm}$:
  - $V(\omega, \bot) = 0$;
  - $V(\omega, F \to G) = 1 \quad \text{iff} \quad V(\omega, F) = 0 \text{ or } V(\omega, G) = 1$;
  - $V(\omega, t : F) = 1 \quad \text{iff} \quad \emptyset \neq \max_{\succcurlyeq} E(\omega, t) \text{ and } \max_{\succcurlyeq}(E(w,t)) \subseteq [F]$

    where $\max_{\succcurlyeq}(S) := \{w \in S \mid (\forall v \in S)(v \succcurlyeq w \to w \succcurlyeq v)\}$.

- E is defined like in $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-subset models with the following conditions for all $w \in W_0$, $s, t \in \mathsf{Tm}^{\mathsf{nm}}$, $A, B \in \mathsf{L}^{\mathsf{nm}}$:

$$\begin{aligned}&E(w, s \cdot t) \neq \emptyset \text{ and}\\ &E(w, s \cdot t) \subseteq \{v \in W \mid \forall F \in \mathsf{APP}_w(s,t)(v \in [F])\}\end{aligned} \tag{7.1}$$

where $\mathsf{APP}_w(s,t)$ is defined analogously to Chapter 2:

$$\begin{aligned}\mathsf{APP}_w(s,t) := \{F \in \mathcal{L}_J^{nm} \mid \exists H \in \mathcal{L}_J^{nm} \text{ s.t.}\\ \max\nolimits_{\succcurlyeq}(E(w,s)) \subseteq [H \to F] \text{ and } \max\nolimits_{\succcurlyeq}(E(\omega,t)) \subseteq [H]\}\end{aligned}$$

$$E(w, s \cup t) = E(w,s) \cup E(w,t) \tag{7.2}$$

$$E(w, s \mathbin{\hat{+}} t) = \max\nolimits_{\succcurlyeq} E(w,s) \cap \max\nolimits_{\succcurlyeq} E(w,t) \tag{7.3}$$

$$E(w, s + t) = E(w,s) \cap E(w,t) \tag{7.4}$$

$$E(w, s \mathbin{\grave{+}} t) = \begin{cases} E(w, s \mathbin{\hat{+}} t) \text{ if } E(w, s \mathbin{\hat{+}} t) \neq \emptyset \\ \max\nolimits_{\succcurlyeq} E(w,s) \text{ otherwise}\end{cases} \tag{7.5}$$

$$\text{if } \mathbf{jd} \in \mathsf{L}^{\mathsf{nm}} : \exists v \in W_0 \text{ with } v \in \max\nolimits_{\succcurlyeq}(E(w,t)) \tag{7.6}$$

$$\text{if } \mathbf{jt} \in \mathsf{L}^{\mathsf{nm}} : w \in \max\nolimits_{\succcurlyeq}(E(w,t)) \tag{7.7}$$

$$\begin{aligned}&\text{if } v \in E(w,s) \text{ then either } v \in \max\nolimits_{\succcurlyeq}(E(w,s)) \text{ or}\\ &\qquad \exists v' \text{ s.t. } v' \succcurlyeq v \text{ and } v' \in \max\nolimits_{\succcurlyeq}(E(w,s))\end{aligned} \tag{7.8}$$

$$\emptyset \neq E(w,c) \subseteq [A] \text{ if } (c,A) \in \mathsf{CS} \tag{7.9}$$

We do not claim antisymmetry for $\succcurlyeq$, i.e. we allow that two worlds or more have the same plausibility and hence it is possible that $w \succcurlyeq v$ and $v \succcurlyeq w$.

In opposition to a tradition in epistemic logic (see for example [13]), we say that $w \succcurlyeq v$ if the world w is more plausible than v. We do this by following the approach of deontic logic as presented by Parent in [29]. The reason to do so is that this approach seems to be more intuitive to us. Nevertheless, it is only a decision on design and does not infect the expressiveness of our framework.

(7.8) is known as smoothness assumption and guarantees that for each world $w \in S$ there is a world $v \in \max_{\succcurlyeq}(S)$ such that $v \succcurlyeq w$. Please note that S may have infinite chains of strictly better worlds, but they do not lead to a single maximal world.[8] It only implies that for each $E(w,s) \neq \emptyset$, the set $\max_{\succcurlyeq}(E(w,s))$ is non-empty. A detailed discussion on smoothness and other limit assumptions can be found in [29].

Definition 7.2 (Truth in $\mathsf{L^{nm}}_{\mathsf{CS}}$-subset models)**.** For an $\mathsf{L^{nm}}_{CS}$-subset model $\mathcal{M} = (W, W_0, \succcurlyeq, V, E)$, a world $\omega \in W$ and a formula F we define the relation $\Vdash$ as follows:

$$\mathcal{M}, \omega \Vdash F \quad \text{iff} \quad V(\omega, F) = 1.$$

7.4. Soundness

Definition 7.3 ($\mathsf{L^{nm}}_{\mathsf{CS}}$-validity)**.** Let CS be a constant specification. We say that a formula $F \in \mathcal{L}_J^{nm}$ is $\mathsf{L^{nm}}_{\mathsf{CS}}$-*valid* if for each $\mathsf{L^{nm}}_{\mathsf{CS}}$-subset model $\mathcal{M} = (W, W_0, \succcurlyeq, V, E)$ and each $\omega \in W_0$, we have $\mathcal{M}, \omega \Vdash F$.

Theorem 7.4 (Soundness of $\mathsf{L^{nm}}_{\mathsf{CS}}$-subset models). *Given a logic $\mathsf{L^{nm}}$ with some constant specification and a formula $F \in \mathsf{L^{nm}}$ we have that if $\mathsf{L^{nm}}_{\mathsf{CS}} \vdash F$, then F is $\mathsf{L^{nm}}_{\mathsf{CS}}$-valid.*

Proof. The proof is by induction on the length of the derivation and analogue to the proof of Theorem 2.4.

We start with the axioms **j**, **jd**, and **jt**:

- If F is an instance of the **j**-axiom:
  Then $F = s : (A \to B) \to (t : A \to s \cdot t : B)$ for some $s, t \in \mathsf{Tm^{nm}}$ and $A, B \in \mathcal{L}_J^{\mathsf{A}}$. Assume for some $w \in W_0$ that $\mathcal{M}, w \Vdash s : (A \to B)$ and $\mathcal{M}, w \Vdash t : A$. Then $V(w, s : (A \to B)) = 1$. Hence since $w \in W_0$ we obtain $\max_{\succcurlyeq}(E(w,s)) \subseteq [A \to B]$ and since $V(w, t : A) = 1$ we obtain $\max_{\succcurlyeq}(E(\omega,t)) \subseteq [A]$. From the definition of $\mathsf{APP}_\omega(r,s)$ we conclude that $B \in \mathsf{APP}_\omega(s,t)$. So for all $v \in E(w, s \cdot t)$ we obtain by the requirements of E that $V(v, B) = 1$ hence $\max_{\succcurlyeq}(E(w, s{\cdot}t)) \subseteq [B]$. But since we defined that $E(w, s \cdot t) \neq \emptyset$ and by smoothness we have that $\emptyset \neq \max_{\succcurlyeq}(E(w, s \cdot t)) \subseteq [B]$. From this, the fact that $w \in W_0$

[8] Consider the following example to see, that infinite chains are possible even with this smoothness condition: We have an infinite number of worlds $\{w_{1.0}, w_{2.0}, \dots, w_{1.1}, w_{2.1} \dots\}$ and $w_{(n+1).0} \succ w_{n.0} \forall n \in \mathbb{N}$. Moreover $w_{n.1} \succ w_{n.0}$. Then there is an infinite chain of strictly better worlds, but the smoothness condition holds.

and the requirements of V in W_0 we obtain $V(w, s \cdot t : B) = 1$, which is $\mathcal{M}, w \Vdash s \cdot t : B$.

- If F is an instance of the **jd**-axiom, then $F = t :\bot \rightarrow \bot$ for some $t \in \mathsf{Tm}^{\mathsf{nm}}$.
  Suppose towards a contradiction that $\mathcal{M}, w \Vdash t :\bot$ for some $t \in \mathsf{Tm}^{\mathsf{nm}}$, then we obtain that $V(\omega, t :\bot) = 1$ and hence by the condition of V in the worlds of W_0, $\max_{\succcurlyeq}(E(\omega, t)) \subseteq [\bot]$. Since $\mathcal{M}$ must be a **jd**-L^{nm}-subset model we claim that $\exists v \in W_0$ s.t. $v \in \max_{\succcurlyeq}(E(w, t))$. But then $v \in [\bot]$ what contradicts that $v \in W_0$.

- If F is an instance of the **jt**-axiom, then $F = t : A \rightarrow A$ for some $A \in \mathsf{L}^{\mathsf{nm}}$ and some $t \in \mathsf{Tm}^{\mathsf{nm}}$.
  Suppose $\mathcal{M}, w \Vdash t : A$. We obtain that $V(w, t : A) = 1$. By the condition on worlds in W_0 we get $\max_{\succcurlyeq}(E(\omega, t)) \subseteq [A]$. Since $\mathcal{M}$ is a **jt**-L^{nm}-subset model, $w \in \max_{\succcurlyeq}(E(\omega, t))$ and therefore we conclude $w \in [A]$. Hence $V(w, A) = 1$ and we obtain that $\mathcal{M}, w \Vdash A$.

- If F is derived by modus ponens the proof is analogue to the one in $\mathsf{L}^{\star}_{\mathsf{CS}}$-subset models.

- If F is derived by axiom necessitation then $F = c : A$ for $(c, A) \in \mathsf{CS}$. Since $(c, A) \in \mathsf{CS}$ we have $\emptyset \neq E(w, c) \subseteq [A]$. Hence with (7.8)

  $$\emptyset \neq \max_{\succcurlyeq}(E(w, c)) \subseteq [A]$$

  and therefore $V(w, c : A) = 1$ and $\mathcal{M}, w \Vdash c : A$.

Before we prove soundness of the remaining axioms, let us have a look how the evidence sets for the new operators are connected together:

$$\max_{\succcurlyeq}(E(w, t)) \subseteq E(w, t) \tag{7.10}$$

$$\max_{\succcurlyeq}(E(w, s \cup t)) \subseteq \max_{\succcurlyeq}(\max_{\succcurlyeq}(E(w, s)) \cup \max_{\succcurlyeq}(E(w, t))) \tag{7.11}$$

$$\max_{\succcurlyeq}(E(w, s \mathbin{\hat{+}} t)) \subseteq \max_{\succcurlyeq}(E(w, s + t)) \tag{7.12}$$

$$\max_{\succcurlyeq}(E(w, s \mathbin{\hat{+}} t)) \subseteq \max_{\succcurlyeq}(E(w, s \cup t)) \tag{7.13}$$

(7.10) is obvious.

(7.11) take some $v \in \max_{\succcurlyeq}(E(w, s \cup t))$. We find that $v \in \max_{\succcurlyeq}(E(w, s))$ or $v \in \max_{\succcurlyeq}(E(w, t))$ and hence $v \in \max_{\succcurlyeq}(E(w, s)) \cup \max_{\succcurlyeq}(E(w, t))$. Now assume towards a contradiction that

$$v \notin \max_{\succcurlyeq}(\max_{\succcurlyeq}(E(w,s)) \cup \max_{\succcurlyeq}(E(w,t))).$$

Hence because of the smoothness condition $\exists v'$ s.t. $v' \succcurlyeq v$ and

$$v' \in \max_{\succcurlyeq}(\max_{\succcurlyeq}(E(w,s)) \cup \max_{\succcurlyeq}(E(w,t))).$$

But

$$\max_{\succcurlyeq}(\max_{\succcurlyeq}(E(w,s)) \cup \max_{\succcurlyeq}(E(w,t))) \subseteq \max_{\succcurlyeq}(E(w,s)) \cup \max_{\succcurlyeq}(E(w,t))$$

and we obtain that $v' \in E(w, s \cup t)$. But then v is not maximal in $E(w, s \cup t)$, which is a contradiction.

(7.12) from $v \in \max_{\succcurlyeq}(E(w, s \hat{+} t))$ we obtain by using (7.10) that $v \in E(w, s+t)$. Assume towards a contradiction that $v \notin \max_{\succcurlyeq}(E(w,s+t))$ and hence because of smoothness $\exists v' \succcurlyeq v$ s.t. $v' \in \max_{\succcurlyeq}(E(w,s+t))$. But then $v' \in E(w,s)$ and $v' \in E(w,t)$ and therefore $v \notin \max_{\succcurlyeq} E(w,s)$ and $v \notin \max_{\succcurlyeq} E(w,t)$ and hence $v \notin \max_{\succcurlyeq} E(w, s \hat{+} t)$, which is a contradiction.

(7.13) Suppose $v \in E(w, s \hat{+} t)$. Hence we have that $v \in \max_{\succcurlyeq}(E(w,s))$ and $v \in \max_{\succcurlyeq}(E(w,t))$. Assume towards a contradiction that

$$v \notin \max_{\succcurlyeq}(E(w, s \cup t)).$$

It is clear that $v \in E(w, s \cup t)$ and hence there must be a v' s.t. $v' \succcurlyeq v$ and $v' \in \max_{\succcurlyeq}(E(w, s \cup t))$. So $v' \in E(w,s)$ or $v' \in E(w,t)$. In the first case we obtain $v \notin \max_{\succcurlyeq} E(w,s)$ and in the second case $v \notin \max_{\succcurlyeq} E(w,t)$, which is a contradiction.

Now we can prove that all instances of the remaining axioms are valid:

- If F is an instance of **NM1**, then $F = s \cup t : A \wedge s \hat{+} t : C \to s \hat{+} t : A$ for some $s, t \in \mathsf{Tm}^{\mathsf{nm}}$ and $A, C \in \mathsf{L}^{\mathsf{nm}}$.

  Suppose $\mathcal{M}, w \Vdash s \cup t : A \wedge s \hat{+} t : C$ then $\max_{\succcurlyeq}(E(w, s \cup t)) \subseteq [A]$. By (7.13) we obtain $\max_{\succcurlyeq}(E(w, s \hat{+} t)) \subseteq [A]$ and since $\mathcal{M}, w \Vdash s \hat{+} t : C$ we have that $V(w, s \hat{+} t : C) = 1$ and hence by using Definition 7.1 we obtain $E(w, s \hat{+} t) \neq \emptyset$. Therefore $V(w, s \hat{+} t : A) = 1$ and hence $\mathcal{M}, w \Vdash s \hat{+} t : A$.

- If F is an instance of **NM2**, then $F = s + t : A \wedge s \hat{+} t : C \to s \hat{+} t : A$ for some $s, t \in \mathsf{Tm}^{\mathsf{nm}}$ and $A, C \in \mathsf{L}^{\mathsf{nm}}$.

Analogue to **NM1** but instead of using (7.13) use (7.12).

- If F is an instance of **NM3**, then $F = s : A \wedge t : A \to s \cup t : A$ for some $s, t \in \mathsf{Tm}^{\mathsf{nm}}$ and $A \in \mathsf{L}^{\mathsf{nm}}$.

  Suppose $\mathcal{M}, w \Vdash s : A \wedge t : A$. We then have $\max_{\succcurlyeq}(E(w, s)) \subseteq [A]$ and $\max_{\succcurlyeq}(E(w, t)) \subseteq [A]$ and hence by (7.11)

  $$\max_{\succcurlyeq}(E(w, s \cup t)) \subseteq \max_{\succcurlyeq}(\max_{\succcurlyeq}(E(w, s)) \cup \max_{\succcurlyeq}(E(w, t))) \subseteq \max_{\succcurlyeq}(E(w, s)) \cup \max_{\succcurlyeq}(E(w, t)) \subseteq [A].$$

  Since $V(w, s : A) = 1$ and by using smoothness, it is clear that

  $$\max_{\succcurlyeq}(E(w, s \cup t)) \neq \emptyset.$$

  Therefore $V(w, s \cup t : A) = 1$ and hence $\mathcal{M}, w \Vdash s \cup t : A$.

- If F is an instance of **NM4**, then

  $$F = (s : A \vee t : A) \wedge s \hat{+} t : C \to s \hat{+} t : A$$

  for some $s, t \in \mathsf{Tm}^{\mathsf{nm}}$ and $A, C \in \mathsf{L}^{\mathsf{nm}}$.

  Suppose $\mathcal{M}, w \Vdash (s : A \vee t : A) \wedge s \hat{+} t : C$ then $\max_{\succcurlyeq}(E(w, s)) \subseteq [A]$ or $\max_{\succcurlyeq}(E(w, t)) \subseteq [A]$ and hence

  $$\max_{\succcurlyeq}(E(w, s)) \cap \max_{\succcurlyeq}(E(w, t)) = E(w, s \hat{+} t) \subseteq [A].$$

  Since $s \hat{+} t : C$ for some $C \in \mathsf{L}^{\mathsf{nm}}$ we obtain that $\max_{\succcurlyeq}(E(w, s \hat{+} t)) \neq \emptyset$. Therefore we obtain with (7.10) that $\max_{\succcurlyeq}(E(w, s \hat{+} t)) \subseteq [A]$ and hence $V(w, s \hat{+} t : A) = 1$ which means $\mathcal{M}, w \Vdash s \hat{+} t : A$.

- If F is an instance of **NM5** then $F = s : A \to s \grave{+} t : A$ for some $s, t \in \mathsf{Tm}^{\mathsf{nm}}$ and $A \in \mathsf{L}^{\mathsf{nm}}$. Suppose moreover $\mathcal{M}, w \Vdash s : A$ then $\max_{\succcurlyeq}(E(w, s)) \subseteq [A]$. Observe that

  $$E(w, s \hat{+} t) = \max_{\succcurlyeq}(E(w, s)) \cap \max_{\succcurlyeq}(E(w, t)) \subseteq \max_{\succcurlyeq}(E(w, s))$$

  and of course

  $$\max_{\succcurlyeq}(E(w, s)) \subseteq \max_{\succcurlyeq}(E(w, s))$$

and hence $E(w, s \dot{+} t) \subseteq [A]$ regardless whether $E(w, s \hat{+} t) = \emptyset$.

- If F is an instance of **NM6**, then

$$F = s \dot{+} t : A \wedge s \hat{+} t : C \to t \dot{+} s : A$$

for some $s, t \in \mathsf{Tm}^{\mathsf{nm}}$ and $A, C \in \mathsf{L}^{\mathsf{nm}}$. Suppose $\mathcal{M}, w \Vdash s \hat{+} t : C$ for some $C \in \mathsf{L}^{\mathsf{nm}}$. This means that $E(w, s \hat{+} t) \neq \emptyset$ and hence $E(w, s \dot{+} t) = E(w, s \hat{+} t)$. But then, by definition, we obtain that $E(w, s \hat{+} t) = E(w, t \hat{+} s)$ and therefore $E(w, t \hat{+} s) \neq \emptyset$ as well and hence $E(w, t \dot{+} s) = E(w, s \dot{+} t)$. Therefore $\max_{\succcurlyeq}(E(w, s \dot{+} t)) \subseteq [A]$ and thus $\mathcal{M}, w \Vdash t \dot{+} s : A$. □

Our axiomatization does not include an axiom where a term $s + t$ occurs on the right hand side of an implication. Indeed, we have the following lemma.

Lemma 7.5. *The formula*

$$s : A \wedge t : A \wedge s \cup t : A \wedge s \dot{+} t : A \wedge s \hat{+} t : A \to s + t : A$$

is not valid.

Proof. Consider Figure 2. We assume that it shows the evidence relations of w_0. Thus $[A] = \{w_1, w_2, w_3\}$, $E(w_0, s) = \{w_0, w_1, w_3\}$, and $E(w_0, t) = \{w_0, w_2, w_3\}$. Hence $\max_{\succcurlyeq}(E(w_0, s)) = \{w_1, w_3\} \subseteq [A]$ and therefore $\mathcal{M}, w_0 \Vdash s : A$, and $\max_{\succcurlyeq}(E(w_0, t)) = \{w_2, w_3\} \subseteq [A]$ thus $\mathcal{M}, w_0 \Vdash t : A$. Furthermore $\max_{\succcurlyeq}(E(w_0, s \cup t)) = \{w_1, w_2, w_3\} \subseteq [A]$ and therefore $\mathcal{M}, w_0 \Vdash s \cup t : A$. Moreover

$$\max_{\succcurlyeq}(E(w_0, s \hat{+} t)) = \max_{\succcurlyeq}(E(w_0, s \dot{+} t)) = \{w_3\} \subseteq [A]$$

and therefore $\mathcal{M}, w_0 \Vdash s \hat{+} t : A \wedge s \dot{+} t : A$. But we also have

$$E(w_0, s + t) = \max_{\succcurlyeq}(E(w_0, s + t)) = \{w_0, w_3\} \not\subseteq [A]$$

and hence $\mathcal{M}, w_0 \nVdash s + t : A$. □

The main reason to extend our set of worlds by adding a preorder on them was to avoid monotonicity. That the application operator is not monotone in $\mathsf{L}^{\mathsf{A}}_{\mathsf{CS}}$-subset models is already shown in Chapter 3. It remains to show that the strategy applied here gives us a non-monotone sum-operator.

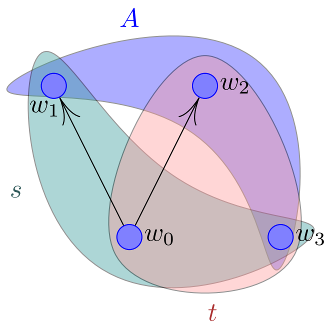

Figure 2.: Example for $s : A \wedge t : A \wedge s \hat{+} t : A \wedge s \grave{+} t : A \wedge s \cup t : A \not\to s + t : A$.

Lemma 7.6 (j+ is not valid in L^{nm}-subset models). *The formula*

$$s : A \vee t : A \to s + t : A$$

is not valid in $\mathsf{L}^{\mathsf{nm}}{}_{\mathsf{CS}}$*-subset models.*

Proof. Follows directly from Lemma 7.5. □

So far we do not have completeness since the standard canonical models we used in the first two sections do not fulfil all the necessary conditions to be $\mathsf{L}^{\mathsf{nm}}{}_{\mathsf{CS}}$-subset models. The problem is not to find a good plausibility relation, but to obtain properties like $E(w, s \cup t) = E(w, s) \cup E(w, t)$. It is quite obvious that some axioms are missing. Hence, this part needs more research.

However, we can show that $\mathsf{L}^{\mathsf{nm}}{}_{\mathsf{CS}}$-subset models exist.

Theorem 7.7 (model existence). *There exists an* $\mathsf{L}^{\mathsf{nm}}{}_{\mathsf{CS}}$*-subset model.*

Proof. We construct a model $\mathcal{M} = (W, W_0, \succcurlyeq, V, E)$ as follows:

- $W = W_0 = \{w\}$
- $w \succcurlyeq w$
- The valuation function is built bottom up:

  - $V(w, \bot) = 0$
  - $V(w, P) = 1$, for all $P \in \mathsf{Prop}$
  - $V(w, A \to B) = 1$ iff $V(w, A) = 0$ or $V(w, B) = 1$
  - $V(w, t : A) = 1$ iff $V(w, A) = 1$

- $E(w, t) = \{w\}$ for all $t \in \mathsf{Tm}^{\mathsf{nm}}$.

By checking all the conditions listed in Definition 7.1 we immediately see that $\mathcal{M}$ is indeed a $\mathsf{L^{nm}}_{\mathsf{CS}}$-subset model. We will only show (7.9):

$$E(w, c) \subseteq [A] \text{ if } (c, A) \in \mathsf{CS}.$$

In fact, we show that all axioms A hold in w and then $E(w, c) = \{w\} = [A]$.

- For all axioms A of **cl** we have $V(w, A) = 1$ by construction of V.

- If A is an instance of **j**, we have $A = s : (F \to G) \to (t : F \to s \cdot t : G)$. Suppose $\mathcal{M}, w \Vdash s : (F \to G) \wedge t : F$ hence $E(w, s) = \{w\} = [F \to G]$ and therefore $V(w, F \to G) = 1$ and similarly $V(w, F) = 1$. By the definition of V we conclude $V(w, G) = 1$. So $E(w, s \cdot t) = \{w\} = [G]$ and hence $V(w, s \cdot t : G) = 1$. Taking everything together leads to

$$V(w, s : (F \to G) \to (t : F \to s \cdot t : G)) = 1$$

  and hence $\{w\} = [A]$.

- If A is an instance of **NM1**, then $A = s \cup t : F \wedge s \hat{+} t : G \to s \hat{+} t : F$. Since $E(w, s \cup t) = E(w, s \hat{+} t) = \{w\}$ we obtain $V(w, s \cup t : F) = 1$ iff $V(w, s \hat{+} t : F) = 1$ and thus $V(w, s \cup t : F \wedge s \hat{+} t : G \to s \hat{+} t : F) = 1$ and therefore $\{w\} = [A]$.

- For all other axioms the proof is analogue to the previous case. $\square$

8. Justification with presumption

When we look around and explore the world, we constantly get new evidence that makes us understand the world we are living in a little bit better. But usually we are not that open to a new piece of evidence that we consider all worlds as possible that do not contradict that specific new piece of evidence. In fact, we have some presumptions about the world, of which we cannot always give explicit justifications, but which we simply added to our belief system at some point in life. We interpret new pieces of evidence in context to these presumptions. In the previous chapters, pieces of evidence have been interpreted as the set of all worlds that are consistent with some observation or information. However, if we receive a new piece of evidence, we often only consider a subset of all those worlds, namely the set of those worlds among them, that are also consistent with our presumptions. So, in this sense we have something like a *selective perception*.

In standard justification logic this is usually not taken into account. With the constant specification we have some very special kind of presumptions, but only about the axioms we believe in. So far, there is no possibility to model that we believe B without indicating the explicit reason why we do so.[9]

The aim of justification logic with presumption is to model an agent's reasoning where not all beliefs are explicitly justified. We do this by allowing justification terms t_Γ where t is a usual term that stands for some evidence and Γ is a set of formulas that are believed without any explicit reason.

[9]There are hybrid justification logics that feature both implicit and explicit knowledge [4]. There, however, the presumptions cannot be reflected on the level of terms.

8.1. Syntax

Definition 8.1 (The language $\mathcal{L}_J^P$)**.** The language $\mathcal{L}_J^P$ is composed from terms and formulas such that

- $0, 1, c_i, x_i$ are atomic terms for constants c_i and variables x_i and the unique constants $0, 1$. All atomic terms are terms.
- If s, t are terms and Γ is a finite non-empty set of formulas, then $s+t$, $s \cdot t, !t$ and t_Γ are terms too.
- $\perp$ and all atomic propositions are formulas.
- if t is a term and F, G are formulas, then $F \to G$ and $t : F$ are formulas too.

We use the abbreviation $t_{A,B}$ to denote the term $t_{\{A,B\}}$.

We extend the logics presented in Chapter 1 by adding new axioms that deal with new kinds of justification.

Definition 8.2 (The logic L^P)**.** Given any logic L^A we define the logic L^P by adding the following axioms to L^A:

P1	$\neg(0 : A)$	for all $A \in \mathcal{L}_J^P$
P2	$1 : A$	for all propositional tautologies A
P3	$t_\Gamma : A$	for all $A \in \Gamma$
P4	$t : A \to t_\Gamma : A$	
P5	$t_A : B \to t + 1 : (A \to B)$	

To deduce formulas in L^P we use a Hilbert system given by L^P and the rules (**MP**) and (**AN!**) as given in chapter 1.

The idea behind **P1** is that 0 is like a blueprint of a piece of evidence so that we can model the presumption without referring to a more detailed justification. So 0_A then is *the* evidence, that A is true.

Justification 1 has a similar function as justification $\mathsf{c}^\star$ in $\mathsf{c}^\star$-subset models but instead of focussing on deductively closed worlds it focuses on normal worlds. We have already used a similar justification for the probabilistic evidence logic in Chapter 6 and which contained **P2** with the same intended interpretation, namely that 1 is *the* justification that all tautologies hold.

P3 claims that if we restrict some evidence to the worlds, where all our presumptions hold, then this is a justification that each of them holds.

The idea behind **P4** is that if we have a justification for something, restricting this justification to the worlds that correspond with our presumptions will not reduce its power to justify a specific formula. So, adding information leads to a monotone process. This may look a bit strange when we consider instances like $t : (\neg A) \to t_A : (\neg A)$ but in fact, this case just illustrates that we are dealing on the one hand with justification that may justify formulas that are evaluated to false (as long as **jt** is not in our logic) and on the other hand with presumptions that may be wrong.

P5 relates the new type of justifications to the standard ones.

8.2. Semantics

Definition 8.3 (L^P-subset model)**.** A model $\mathcal{M} = (W, W_0, E, V)$ is called an L^P-subset model, if W is a set of worlds that contains a particular world $w_\emptyset$, and W_0, V and E are defined analogue to Definition 2.1 with the following condition added on V and E for all $w \in W_0$, $t, t_\Gamma \in \mathsf{Tm}$, $A \in \mathcal{L}_J^P$:

- $V(\omega_\emptyset, F) = 0, \quad \forall F \in \mathcal{L}_J^P$;
- $E(\omega, t_\Gamma) = \left(\bigcap_{A\in\Gamma}[A]\right) \cap E(\omega, t)$;
- $E(\omega, 0) = W$;
- $E(\omega, 1) = W_0$.

This new world $\omega_\emptyset$ is of course not an element of W_0 and models a world where nothing at all is true. We have to add it in order to be sure that 0 by itself does not justify anything.

Truth in an L^P-subset models is defined as before, i.e.

$$\mathcal{M}, \omega \Vdash F \quad \text{iff} \quad V(\omega, F) = 1.$$

8.3. Soundness

Definition 8.4 (L^P-validity)**.** Let CS be a constant specification. We say that a formula $F \in \mathcal{L}_J^P$ is ${\mathsf{L}^\mathsf{P}}_\mathsf{CS}$-*valid* if for each ${\mathsf{L}^\mathsf{P}}_\mathsf{CS}$-subset model $\mathcal{M} = (W, W_0, V, E)$ and each $w \in W_0$, we have $\mathcal{M}, w \Vdash F$.

Theorem 8.5 (Soundness of $\mathsf{L^P}_{\mathsf{CS}}$-subset models). *For any justification logic with presumptions $\mathsf{L^P}$, any constant specification CS and any formula $F \in \mathcal{L}_J^P$ we have that if $\mathsf{L^P}_{\mathsf{CS}} \vdash F$, then F is $\mathsf{L^P}_{\mathsf{CS}}$-valid.*

Proof. The proof is again by induction on the length of the derivation and analogue to the proof of Theorem 2.4. Since $\mathsf{L^P}$-subset models only differ in the aspects of these new axioms, we only show this part of the proof here.

- If F is an instance of **P1** then $F = \neg(0 : A)$ for some formula A.
  Since $\omega_\emptyset \in W$ we obtain that $[A] \subsetneq W$ and hence $W = E(\omega, 0) \not\subseteq [A]$ for all formulas A and $\omega \in W_0$. So $V(\omega, 0 : A) = 0$ and hence $\mathcal{M}, \omega \not\Vdash 0 : A$ and therefore $\mathcal{M}, \omega \Vdash \neg(0 : A)$.

- If F is an instance of **P2** then $F = 1 : A$ for some tautology A.
  Since for all $\omega \in W_0$ the valuation V is defined such that tautologies are evaluated to 1 we obtain that $W_0 \subseteq [A]$ for every tautology A. And so we conclude $W_0 = E(\omega, 1) \subseteq [A]$ for all $\omega \in W_0$ and hence $V(\omega, 1 : A) = 1$ and finally $\mathcal{M}, \omega \Vdash 1 : A$.

- If F is an instance of **P3** then $F = t_\Gamma : A$ for some justification t, some set of formulas Γ and some $A \in \Gamma$.
  Since $A \in \Gamma$ we obtain $\bigcap_{B \in \Gamma} [B] \subseteq [A]$ and then any further intersection on the left side is of course as well a subset of $[A]$. Therefore

  $$E(\omega, t_\Gamma) = \left(\bigcap_{B \in \Gamma} [B] \right) \cap E(\omega, t) \subseteq [A]$$

  and hence $V(\omega, t_\Gamma : A) = 1$ and finally $\mathcal{M}, \omega \Vdash t_\Gamma : A$.

- If F is an instance of **P4** then $F = t : A \to t_\Gamma : A$ for some $t \in \mathsf{Tm}$, $A \in \mathcal{L}_J^P$ and some set of formulas Γ.
  Suppose $\mathcal{M}, \omega \Vdash t : A$ then $E(\omega, t) \subseteq [A]$. Since

  $$E(\omega, t_\Gamma) = \left(\bigcap_{B \in \Gamma} [B] \right) \cap E(\omega, t) \subseteq E(\omega, t) \subseteq [A]$$

  we obtain $V(\omega, t_\Gamma : A) = 1$ and conclude $\mathcal{M}, \omega \Vdash t_\Gamma : A$.

- If F is an instance of **P5** then $F = t_A : B \to t + 1 : (A \to B)$ for some term t and formulas A, B.
  Suppose $\mathcal{M}, \omega \Vdash t_A : B$ then $V(\omega, t_A : B) = 1$ and therefore
  $$E(\omega, t) \cap [A] \subseteq [B] \tag{8.1}$$
  If $E(\omega, t+1) = \emptyset$ then we directly obtain $E(\omega, t+1) \subseteq [A \to B]$ and hence $V(\omega, t+1 : (A \to B)) = 1$ and thus $\mathcal{M}, \omega \Vdash t+1 : (A \to B)$.
  If $E(\omega, t+1) \neq \emptyset$, take some arbitrary $\upsilon \in E(\omega, t+1) \subseteq E(\omega, t) \cap W_0$. We have either $\upsilon \in [A]$ or $\upsilon \notin [A]$.
  - If $\upsilon \in [A]$ we obtain by $\upsilon \in E(w, t)$ and (8.1) that $\upsilon \in [B]$. The conditions on W_0 further allow us to conclude from $V(\upsilon, B) = 1$ that $V(\upsilon, A \to B) = 1$ and hence $\upsilon \in [A \to B]$.
  - If $\upsilon \notin [A]$ then we can directly deduce from $V(\upsilon, A) = 0$ and $\upsilon \in W_0$ that $V(\upsilon, A \to B) = 1$ and hence $\upsilon \in [A \to B]$.

  So both $\upsilon \in [A]$ and $\upsilon \notin [A]$ imply $\upsilon \in [A \to B]$. Therefore we find $E(\omega, t+1) \subseteq [A \to B]$ and conclude $\mathcal{M}, \omega \Vdash t+1 : (A \to B)$. □

Remark 8.6. So far, we have no completeness. Following the strategy we had in Chapters 1 and 2 will not work here. It would be possible to define both W^C and W_0^C in the canonical model. For the latter we would have to define something like extended maximally $\mathsf{L}^{\star}_{\mathsf{CS}}$-consistent sets of formulas, where we would use an algorithm similar to the one of Lindenbaum to decide whether to add formulas with the new defined justification terms or not to some maximally $\mathsf{L}^{\star}_{\mathsf{CS}}$-consistent set of formulas. However, the definition of E^C in such a canonical modal will not lead neither to $E^C(\Gamma, 0) = W^C$ nor $E^C(\Gamma, 1) = W_0^C$.

However, we can show that L^{P}-models exist.

Before we start to define such a model, we make some definitions that will be useful later.

Definition 8.7 ($\mathfrak{p}_{\text{term}}$)**.** $\mathfrak{p}_{\text{term}}$'s are defined inductively as follows:

- 0 and 1 are $\mathfrak{p}_{\text{term}}$'s.
- If Γ is a non-empty finite set of $\mathcal{L}_J^P$-formulas and t is a $\mathfrak{p}_{\text{term}}$ then t_Γ is a $\mathfrak{p}_{\text{term}}$.
- If s and t are $\mathfrak{p}_{\text{term}}$'s, then $s + t$ is a $\mathfrak{p}_{\text{term}}$.

So basically $\mathfrak{p}_{\text{term}}$'s are justification terms that have no other variables or constants than 0 and 1 and do not contain any !'s.

Definition 8.8 ($\mathfrak{p}_{\text{set}}$)**.** Given a $\mathfrak{p}_{\text{term}}$ t we define the set of formulas $\mathfrak{p}_{\text{set}}$ as follows:

$$\mathfrak{p}_{\text{set}}(t) := \{F \in \mathcal{L}_J^P \mid F \in \Gamma \text{ for some } \Gamma \text{ occuring in the subscript of } t\}$$

Definition 8.9. We define the depth of a term respectively of a formula δ inductively as follows:

$$\begin{aligned}
&\delta(s) := 0 \text{ if } s \in \mathsf{ATm}\\
&\delta(!s) := \delta(s)\\
&\delta(s+t) := \max(\delta(s), \delta(t)) + 1\\
&\delta(s_\Gamma) := \delta(s) + \max(\delta(A) \text{ for } A \in \Gamma) + 1\\
&\delta(\bot) := 0\\
&\delta(P) := 0 \text{ for } P \in \mathsf{Prop}\\
&\delta(A \to B) := \max(\delta(A), \delta(B)) + 1\\
&\delta(s : A) := \max(\delta(s), \delta(A)) + 1
\end{aligned}$$

Theorem 8.10 (model existence). *There exists an* L^P*-subset model.*

Proof. We show this for the logic $\mathsf{L}^\star$ with empty $\mathsf{L}^\star_\beta$ and an empty CS and construct a model $\mathcal{M} = (W, W_0, V, E)$ as follows:

- $W = \{w_\emptyset, w\}$
- $W_0 = \{w\}$
- $V(w_\emptyset, F) = 0 \quad \forall F \in \mathcal{L}_J^P$
  The valuation function F in w is defined by induction on $\delta(F)$ bottom up:
  - $V(w, \bot) = 0$
  - $V(w, P) = 1 \quad \forall P \in \mathsf{Prop}$
  - $V(w, A \to B) = 1 \quad$ iff $\quad V(w, A) = 0$ or $V(w, B) = 1$

– $V(w, t : A) = \begin{cases} 0 & \text{if } t = 0 \text{ or } (t \neq 0 \text{ is a } \mathfrak{p}_{\text{term}}, V(w, A) = 0 \\ & \text{and } \forall B \in \mathfrak{p}_{\text{set}}(t)(V(w, B) = 1)) \\ 1 & \text{otherwise.} \end{cases}$

- E is defined in parallel to $V(w, F)$ as well by induction on δ as follows:
  – $E(w_\emptyset, t) = W \quad \forall t \in \mathsf{Tm}$.
  – $E(w, t) = \begin{cases} W & \text{if } t = 0 \\ W_0 & \text{if } t = 1 \\ \bigcap_{A \in \mathfrak{p}_{\text{set}}(t)} [A] & \text{for all other } \mathfrak{p}_{\text{term}}\text{'s} \\ \emptyset & \text{otherwise} \end{cases}$

In order to prove that $\mathcal{M}$ is indeed an L^{P}-subset model we have to show that it fulfils all conditions listed in Definition 1.2 and Definition 8.3. Most of them follow directly from the definition of $\mathcal{M}$ or are very easy to see. We only show that $V(w, t : F) = 1$ iff $E(w, t) \subseteq [F]$. We have to distinguish the following cases:

(1) $t = 1$: Here we have two subcaes:
   a) $V(w, F) = 0$: then we have $V(w, t : F) = 0$ by definition of V in $\mathcal{M}$. Furthermore $E(w, t) = W_0$ and $[F] = \emptyset$ and hence $E(w, t) \not\subseteq [F]$.
   b) $V(w, F) = 1$: then $[F] = W_0$ and hence $E(w, t) \subseteq [F]$.

(2) $t = 0$: Since $E(w, 0) = W$ and $w_\emptyset \in W$ we obtain that $E(w, 0) \not\subseteq [F]$. By definition of V there is $V(w, 0 : F) = 0$.

(3) $t \neq 0$ is a $\mathfrak{p}_{\text{term}}$: Hence it is easy to see that $E(w, t) = \bigcap_{A \in \mathfrak{p}_{\text{set}}(t)} [A]$. We distinguish two main cases:
   a) $\exists A \in \mathfrak{p}_{\text{set}}(t)$ with $V(w, A) = 0$. Then $E(w, t) = \emptyset$ and hence $E(w, t) \subseteq [F]$ anyway which corresponds to $V(w, t : F) = 1$.
   b) $\forall A \in \mathfrak{p}_{\text{set}}(t)$ there is $V(w, A) = 1$. Hence $E(w, t) = \{w\}$. We have again to distinguish two cases:
      i. $V(w, F) = 1$: then we obtain directly $E(w, t) \subseteq [F]$ which corresponds to the fact that $V(w, t : F) = 1$.
      ii. $V(w, F) = 0$: hence $V(w, t : F) = 0$ which corresponds to $E(w, t) = \{w\} \not\subseteq [F] = \emptyset$.

(4) If t is something else: since then $E(w,t) = \emptyset$ we have $E(w,t) \subseteq [F]$ anyway which corresponds to $V(w, t : F) = 1$. $\square$

9. Contraction

In belief revision contraction refers to the operation of 'removing' a sentence from a belief set. Justification logic is normally used to model how we obtain new justifications by combining older ones. In this sense it is quite strange to have a section on contraction, which, in fact, is a process where we lose information. Nevertheless, we tried to find a way of modelling contraction within models of justification logic.

We model contraction by the following two formulas:

$$s : A \;\rightarrow\; s^{-B} : A \quad \text{for } A \neq B, \tag{9.1}$$

$$\neg(s^{-B} : B), \tag{9.2}$$

where s^{-B} denotes the justification s that loses the capacity to justify B. In other words: if justification s has the capacity to justify B, then the set of worlds within the interpretation of s is changed in s^{-B} such that it does no longer justify B. So (9.1) guarantees that s^{-B} only loses its power to justify B but apart from this, the justification keeps its power to justify all other formulas that s justified. And (9.2) guarantees that B no longer is justified by s^{-B}.

To model these new features within subset models we have to adapt our syntax and semantics.

Definition 9.1 (The language $\mathcal{L}_J^C$)**.** The language $\mathcal{L}_J^C$ is composed from terms and formulas such that:

- c_i, x_i, are atomic terms for constants c_i and variables x_i. All atomic terms are terms.
- If s, t are terms and B is a formula, then $t + t$, $s \cdot t$, $!t$ and t^{-B} are terms too.
- $\perp$ and all atomic propositions are formulas.

- if t is a term and F, G are formulas, then $F \to G$ and $t : F$ are formulas too.

Definition 9.2 (The logic L^C)**.** Given a logic L^A we define the logic L^C by adding the following two new axioms to L^A:

C1 $\quad s : A \ \to \ s^{-B} : A \quad$ for $A \neq B$
C2 $\quad \neg(s^{-B} : B)$

Definition 9.3 (Non-B world)**.** Given a model $\mathcal{M} = (W, W_0, E, V)$ and a formula B, we say $\omega \in W$ is a *non-B world* if $V(\omega, B) = 0$. Further we say $\omega \in W$ is a *maximal non-B world* if

$$V(\omega, A) = \begin{cases} 0 & \text{if } A = B \\ 1 & \text{otherwise.} \end{cases}$$

Obviously for any $B \in \mathcal{L}_J^C$, a maximal non-B world ω is not consistent with classical logic. Since we allow non-normal worlds in our models, this is not a problem. In general, there is not a unique non-B world for some formula B, since worlds may have the same valuations but differ in their evidence function.

Definition 9.4 (L^C-subset model)**.** A model $\mathcal{M} = (W, W_0, E, V)$ is called an $\mathcal{L}_J^C$-subset model, if W, W_0, E, V are defined analogously to Definition 2.1 and

- for each formula B, there exists at least one maximal non-B world. For each formula B we pick one such maximal non-B-world and denote it with $\omega_{\overline{B}}$.
- E additionally satisfies for all $\omega \in W_0$ and all $s^{-B} \in \mathsf{Tm}$:

$$E(\omega, s^{-B}) = E(\omega, s) \cup \{\omega_{\overline{B}}\}.$$

Truth in an L^C-subset models is defined as before, i.e.

$$\mathcal{M}, \omega \Vdash F \quad \text{iff} \quad V(\omega, F) = 1.$$

Theorem 9.5 (Soundness). *Given a logic L^C and a formula F*

$$\mathsf{L}^\mathsf{C} \vdash F \Rightarrow \mathcal{M}, \omega \Vdash F \textit{ for all } \omega \in W_0 \textit{ in all } \mathsf{L}^\mathsf{C}\textit{-subset models } \mathcal{M}$$

Proof. The proof is analogue to the proof of Theorem 2.4. We only show the cases for the additional axioms.

- If F is an instance of **C1** then $F = s : A \to s^{-B} : A$ for $A \neq B$. Suppose $\mathcal{M}, \omega \Vdash s : A$ i.e. $E(\omega, s) \subseteq [A]$ for $\omega \in W_0$. Since $A \neq B$, we know $\omega_{\overline{B}} \in [A]$ and therefore $E(\omega, s^{-B}) = E(\omega, s) \cup \{\omega_{\overline{B}}\} \subseteq [A]$. We conclude $\mathcal{M}, \omega \Vdash s^{-B} : A$.

- If F is an instance of **C2** then $F = \neg(s^{-B} : B)$ for some $B \in \mathcal{L}_J^C$. Since $\omega_{\overline{B}} \notin [B]$, it is obvious that $E(\omega, s^{-B}) = E(\omega, s) \cup \{\omega_{\overline{B}}\} \not\subseteq [B]$. Hence $\mathcal{M}, \omega \nVdash s^{-B} : B$. $\square$

Remember that we work in a hyperintensional context. Applying the contraction operator $\cdot^{-B}$ to a term s only removes B from the formulas justified by s. In particular, we may have that $s^{-B} : (A \wedge B)$ is true although $s^{-B} : B$ must be false. This could be addressed by introducing some kind of selection function that chooses a (non-maximal) non-B set that satisfies certain closure conditions to define the interpretation of $\cdot^{-B}$. Of course then axiom **C1** needs to be changed accordingly.

10. Updates

In the context of justification logic an update is seen as a belief change operator. If we interpret justifications as sets of formulas, as it is done in most semantics, such a belief change operator changes these sets. Such dynamic epistemic justification logics have been studied, e.g. in [15, 16, 22, 32]. Kuznets and Studer [22], in particular, introduce a justification logic with an operation for belief expansion. Their system satisfies a Ramsey principle as well as minimal change. In fact, their system meets all AGM postulates for belief expansion.

In their model, the belief expansion operation is monotone: belief sets can only get larger, i.e.,

$$\text{belief expansion always only adds new beliefs.} \tag{10.1}$$

This is fine for first-order beliefs. Indeed, one of the AGM postulates for expansion requires that beliefs are persistent. However, as we will argue later, this behavior is problematic for higher-order beliefs.

In this Chapter, we present an alternative approach that behaves better with respect to higher-order beliefs and uses subset models for justification logics.

It is the aim of this Chapter to equip subset models with an operation for belief expansion similar to [22]. The main idea is to introduce justification terms $\mathsf{up}(A)$ such that after a belief expansion with A, we have that A is believed and $\mathsf{up}(A)$ (representing the expansion operation on the level of terms) justifies this belief. Semantically, the expansion with A is dealt with by intersecting the interpretation of $\mathsf{up}(A)$ with the truth-set of A. This provides a better approach to belief expansion than [22] as (10.1) will hold for first-order beliefs but it will fail in general.

The Chapter is organized as follows. First, we introduce the language and a deductive system for JUS, a justification logic with belief expansion and subset models. Then we present its semantics and establish soundness of JUS. Section 10.4 is concerned with persistence properties of first-order and higher-order beliefs. Finally we prove a Ramsey property for JUS.

10.1. Syntax

Given a set of countably many constants c_i, countably many variables x_i, and countably many atomic propositions P_i, terms and formulas of the language of JUS are defined as follows:

- Evidence terms
  - Every constant c_i and every variable x_i is an atomic term. If A is a formula, then $\mathsf{up}(A)$ is an atomic term. Every atomic term is a term.
  - If s and t are terms and A is a formula, then $s \cdot_A t$ is a term.
- Formulas
  - Every atomic proposition P_i is a formula.
  - If A, B, C are formulas, and t is a term, then $\neg A$, $A \to B$, $t : A$ and $[C]A$ are formulas.

The annotation of the application operator may seem a bit odd at first. However, it is often used in dynamic epistemic justification logics, see, e.g. [22].

The set of atomic terms is denoted by ATm, the set of all terms is denoted by Tm. The set of atomic propositions is denoted by Prop and the set of all formulas is denoted by $\mathsf{L}_{\mathsf{JUS}}$. We define the remaining classical connectives, $\bot$, $\wedge$, $\vee$, and $\leftrightarrow$, as usual making use of the law of double negation and de Morgan's laws.

The intended meaning of the justification term $\mathsf{up}(A)$ is that after an update with A, this act of updating serves as justification to believe A. Consequently, the justification term $\mathsf{up}(A)$ has no specific meaning before the update with A happens.

Definition 10.1 (Set of Atomic Subterms)**.** The set of atomic subterms of a term or formula is inductively defined as follows:

- $\mathsf{atm}(t) := \{t\}$ if t is a constant or a variable
- $\mathsf{atm}(\mathsf{up}(C)) := \{\mathsf{up}(C)\} \cup \mathsf{atm}(C)$

- $\mathsf{atm}(s \cdot_A t) := \mathsf{atm}(s) \cup \mathsf{atm}(t) \cup \mathsf{atm}(A)$
- $\mathsf{atm}(P) := \emptyset \quad$ for $P \in \mathsf{Prop}$
- $\mathsf{atm}(\neg A) := \mathsf{atm}(A)$
- $\mathsf{atm}(A \to B) := \mathsf{atm}(A) \cup \mathsf{atm}(B)$
- $\mathsf{atm}(t : A) := \mathsf{atm}(t) \cup \mathsf{atm}(A)$
- $\mathsf{atm}([C]A) := \mathsf{atm}(A) \cup \mathsf{atm}(C)$.

Definition 10.2. We call a formula A *up-independent* if for each subformula $[C]B$ of A we have that $\mathsf{up}(C) \notin \mathsf{atm}(B)$.

Using Definition 10.1, we can control that updates and justifications are independent. This is of importance to distinguish cases where updates change the meaning of justifications and corresponding formulas from cases where the update does not affect the meaning of a formula.

We will use the following notation: τ denotes a finite sequence of formulas and ϵ denotes the empty sequence. Given a sequence $\tau = C_1, \dots C_n$ and a formula A, the formula $[\tau]A$ is defined by

$$[\tau]A = [C_1] \dots [C_n]A \text{ if } n > 0 \quad \text{and} \quad [\epsilon]A := A.$$

The logic JUS has the following axioms and rules where τ is a finite (possibly empty) sequence of formulas:

1. $[\tau]A \quad$ for all propositional tautologies A (Taut)
2. $[\tau](t : (A \to B) \land s : A \leftrightarrow t \cdot_A s : B)$ (App)
3. $[\tau]([C]A \leftrightarrow A) \quad$ if $[C]A$ is up-independent (Indep)
4. $[\tau]([C]\neg A \leftrightarrow \neg[C]A)$ (Funct)
5. $[\tau]([C](A \to B) \leftrightarrow ([C]A \to [C]B))$ (Norm)
6. $[\tau][A]\mathsf{up}(A) : A$ (Up)
7. $[\tau](\mathsf{up}(A) : B \to [A]\mathsf{up}(A) : B)$ (Pers)

A constant specification CS for JUS is any subset

$$\begin{aligned}\mathsf{CS} \subseteq \{(c,\ [\tau_1]c_1 : [\tau_2]c_2 : \ldots : [\tau_n]c_n : A) \ \mid \\ & n \geq 0,\ c, c_1, \ldots, c_n \text{ are constants,} \\ & \tau_1, \ldots, \tau_n \text{ are sequences of formulas,} \\ & A \text{ is an axiom of } \mathsf{JUS}\}\end{aligned}$$

$\mathsf{JUS}_{\mathsf{CS}}$ denotes the logic JUS with the constant specification CS. The rules of $\mathsf{JUS}_{\mathsf{CS}}$ are Modus Ponens and Axiom Necessitation:

$$\frac{A \qquad A \to B}{B}\ (\mathrm{MP}) \qquad\qquad \frac{}{[\tau]c : A}\ (\mathrm{AN}) \quad \text{if } (c, A) \in \mathsf{CS}$$

Before establishing some basic properties of $\mathsf{JUS}_{\mathsf{CS}}$, let us briefly discuss its axioms. The direction from left to right in axiom (App) provides an internalization of modus ponens. Because of the annotated application operator, we also have the other direction, which is a minimality condition. It states that a justification represented by a complex term can only come from an application of modus ponens.

Axiom (Indep) roughly states that an update with a formula C can only affect the truth of formulas that contain certain update terms.

Axiom (Funct) formalizes that updates are functional, i.e. the result of an update is uniquely determined.

Axiom (Norm), together with Lemma 10.4, states that $[C]$ is a normal modal operator for each formula C.

Axiom (Up) states that after a belief expansion with A, the formula A is indeed believed and $\mathsf{up}(A)$ justifies that belief.

Axiom (Pers) is a simple persistency property of update terms.

Definition 10.3. A constant specification CS is called *axiomatically appropriate* if

(1) for each axiom A, there is a constant c with $(c, A) \in \mathsf{CS}$ and

(2) for any formula A and any constant c, if $(c, A) \in \mathsf{CS}$, then for each sequence of formulas τ there exists a constant d with

$$(d, [\tau]c : A) \in \mathsf{CS}.$$

The first clause in the previous definition is the usual condition for an axiomatically appropriate constant specification (when the language includes the !-operation). Here we also need the second clause in order to have the following two lemmas which establish that necessitation is admissible in $\mathsf{JUS}_{\mathsf{CS}}$. Both are proved by induction on the length of derivations.

Lemma 10.4. *Let* CS *be an arbitrary constant specification. For all formulas A and C we have that if A is provable in* $\mathsf{JUS}_{\mathsf{CS}}$*, then $[C]A$ is provable in* $\mathsf{JUS}_{\mathsf{CS}}$*.*

Lemma 10.5 (Constructive Necessitation). *Let* CS *be an axiomatically appriopriate constant specification. For all formulas A we have that if A is provable in* $\mathsf{JUS}_{\mathsf{CS}}$*, then there exists a term t such that $t : A$ is provable in* $\mathsf{JUS}_{\mathsf{CS}}$*.*

We will also need the following auxiliary lemma.

Lemma 10.6. *Let* CS *be an arbitrary constant specification. For all terms s, t and all formulas A, B, C,* $\mathsf{JUS}_{\mathsf{CS}}$ *proves:*

$$[C]t : (A \to B) \wedge [C]s : A \ \leftrightarrow \ [C]t \cdot_A s : B$$

10.2. Semantics

Now we are going to introduce subset models for the logic $\mathsf{JUS}_{\mathsf{CS}}$. In order to define a valuation function on these models, we will need the following measure for the length of formulas.

Definition 10.7 (Length)**.** The *length* of a term or formula is inductively defined by:

$$\begin{array}{ll}
\ell(t) := 1 \text{ if } t \in \mathsf{ATm} & \ell(s \cdot_A t) := \ell(s) + \ell(t) + \ell(A) + 1 \\
\ell(P) := 1 \text{ if } P \in \mathsf{Prop} & \ell(A \to B) := \ell(A) + \ell(B) + 1 \\
\ell(\neg A) := \ell(A) + 1 & \ell(t : A) := \ell(t) + \ell(A) + 1 \\
\ell([B]A) := \ell(B) + \ell(A) + 1 &
\end{array}$$

Definition 10.8 (Subset Model)**.** We define a *subset model*

$$\mathcal{M} = (W, W_0, V_1, V_0, E)$$

for JUS by:

- W is a set of objects called worlds.
- $W_0 \subseteq W$, $W_0 \neq \emptyset$.
- $V_1 : (W \setminus W_0) \times \mathsf{L_{JUS}} \to \{0,1\}$.
- $V_0 : W_0 \times \mathsf{Prop} \to \{0,1\}$.
- $E : W \times \mathsf{Tm} \to \mathcal{P}(W)$ such that for $\omega \in W_0$ and all $A \in \mathsf{L_{JUS}}$:

  $$E(\omega, s \cdot_A t) \subseteq E(\omega, s) \cap E(\omega, t) \cap W_{MP},$$

  where W_{MP} is the set of all deductively closed worlds, formally given by

  $$\begin{aligned} W_{MP} &:= W_0 \cup W^1_{MP} \quad \text{where} \\ W^1_{MP} &:= \{\omega \in W \setminus W_0 \mid \\ &\qquad \forall A, B \in \mathsf{L_{JUS}} \; ((V_1(\omega, A) = 1 \text{ and } V_1(\omega, A \to B) = 1) \\ &\qquad\qquad \text{implies } V_1(\omega, B) = 1)\}. \end{aligned}$$

As in the previous chapters we call W_0 the set of *normal* worlds. The worlds in $W \setminus W_0$ are called *non-normal* worlds. W_{MP} denotes the set of worlds where the valuation function (see the following definition) is closed under modus ponens.

Let $\mathcal{M} = (W, W_0, V_1, V_0, E)$ be a subset model. We define the *valuation function* $V_{\mathcal{M}}$ for $\mathcal{M}$ and the *updated model* $\mathcal{M}^C$ for any formula C simultaneously. For $V_{\mathcal{M}}$, we often drop the subscript $\mathcal{M}$ if it is clear from the context.

We define $V : W \times \mathsf{L_{JUS}} \to \{0,1\}$ as follows by induction on the length of formulas:

(1) Case $\omega \in W \setminus W_0$. We set $V(\omega, F) := V_1(\omega, F)$;

(2) Case $\omega \in W_0$. We define V inductively by:

   a) $V(\omega, P) := V_0(\omega, P)$ for $P \in \mathsf{Prop}$;

   b) $V(\omega, t : F) := 1$ iff $E(\omega, t) \subseteq \{\upsilon \in W \mid V(\omega, F) = 1\}$ for $t \in \mathsf{ATm}$;

c) $V(\omega, s \cdot_F r : G) = 1$ iff
$V(\omega, s : (F \to G)) = 1$ and $V(\omega, r : F) = 1$;

d) $V(\omega, \neg F) = 1$ iff $V(\omega, F) = 0$;

e) $V(\omega, F \to G) = 1$ iff $V(\omega, F) = 0$ or $V(\omega, G) = 1$;

f) $V(\omega, [C]F) = 1$ iff $V_{\mathcal{M}^C}(\omega, F) = 1$ where $V_{\mathcal{M}^C}$ is the valuation function for the updated model $\mathcal{M}^C$.

The following notation for the truth set of F will be convenient:

$$[[F]]_{\mathcal{M}} := \{v \in W \mid V_{\mathcal{M}}(v, F) = 1\}.$$

The updated model $\mathcal{M}^C = (W^{\mathcal{M}^C}, W_0^{\mathcal{M}^C}, V_1^{\mathcal{M}^C}, V_0^{\mathcal{M}^C}, E^{\mathcal{M}^C})$ is given by:

$$W^{\mathcal{M}^C} := W \qquad W_0^{\mathcal{M}^C} := W_0 \qquad V_1^{\mathcal{M}^C} := V_1 \qquad V_0^{\mathcal{M}^C} := V_0$$

and

$$E^{\mathcal{M}^C}(\omega, t) := \begin{cases} E^{\mathcal{M}}(\omega, t) \cap [[C]]_{\mathcal{M}^C} & \text{if } \omega \in W_0 \text{ and } t = \mathsf{up}(C) \\ E^{\mathcal{M}}(\omega, t) & \text{otherwise} \end{cases}$$

The valuation function for complex terms is well-defined.

Lemma 10.9. *For a subset model $\mathcal{M}$ with a world $\omega \in W_0$, $s, t \in \mathsf{Tm}$, $A, B \in \mathsf{L_{JUS}}$, we find that*

$$V(\omega, s \cdot_A t : B) = 1 \quad \textit{implies} \quad E(\omega, s \cdot_A t) \subseteq [[B]]_{\mathcal{M}}.$$

Proof. The proof is by induction on the structure of s and t:

- base case $s, t \in \mathsf{ATm}$:
  Suppose $V(\omega, s \cdot_A t : B) = 1$. Case 2c of the definition of V in Definition 10.8 for normal worlds yields that

  $$V(\omega, s : (A \to B)) = 1 \text{ and } V(\omega, t : A) = 1.$$

  With case 2b from the same definition we obtain

  $$E(\omega, s) \subseteq [[A \to B]]_{\mathcal{M}} \text{ and } E(\omega, t) \subseteq [[A]]_{\mathcal{M}}.$$

Furthermore the definition of E for normal worlds guarantees that

$$E(\omega, s \cdot_A t) \subseteq E(\omega, s) \cap E(\omega, t) \cap W_{MP}.$$

So for each $\upsilon \in E(\omega, s \cdot_A t)$ there is $V(\upsilon, A \to B) = 1$ and $V(\upsilon, A) = 1$ and $\upsilon \in W_{MP}$ and hence either by the definition of W^1_{MP} or by case 2e of the definition of V in normal worlds there is $V(\upsilon, B) = 1$. Therefore $E(\omega, s \cdot_A t) \subseteq [[B]]_{\mathcal{M}}$.

- $s, t \in \mathsf{Tm}$ but at least one of them is not atomic: w.l.o.g. suppose $s = r \cdot_C q$. Suppose $V(\omega, s \cdot_A t : B) = 1$ then $V(\omega, s : (A \to B)) = 1$ and $V(\omega, t : A) = 1$. Since $s = r \cdot_C q$ and $\omega \in W_0$ we obtain

  $$V(\omega, r : (C \to (A \to B))) = 1 \text{ and } V(\omega, q : C) = 1$$

  and by I.H. that

  $$E(\omega, r) \subseteq [[C \to (A \to B)]]_{\mathcal{M}} \text{ and } E(\omega, q) \subseteq [[C]]_{\mathcal{M}}.$$

  With the same reasoning as in the base case we obtain

  $$E(\omega, s) = E(\omega, r \cdot_C q) \subseteq [[A \to B]]_{\mathcal{M}}.$$

  If t is neither atomic, the argumentation works analogue and since we have then shown both $E(\omega, s) \subseteq [[A \to B]]_{\mathcal{M}}$ and $E(\omega, t) \subseteq [[A]]_{\mathcal{M}}$, the conclusion is the same as in the base case. □

Remark 10.10. The opposite direction to Lemma 10.9 need not hold. Consider a model $\mathcal{M}$ and a formula $s \cdot_A t : B$ with atomic terms s and t such that

$$V_{\mathcal{M}}(\omega, s \cdot_A t : B) = 1$$

and thus also $E(\omega, s \cdot_A t) \subseteq [[B]]_{\mathcal{M}}$. Now consider a model $\mathcal{M}'$ which is defined like $\mathcal{M}$ except that

$$E'(\omega, s) := E(\omega, t) \quad \text{and} \quad E'(\omega, t) := E(\omega, s).$$

We observe the following:

(1) We have $E'(\omega, s \cdot_A t) = E(\omega, s \cdot_A t)$ as the condition

$$E'(\omega, s \cdot_A t) \subseteq E'(\omega, s) \cap E'(\omega, t) \cap W_{MP}$$

still holds since intersection of sets is commutative. Therefore

$$E'(\omega, s \cdot_A t) \subseteq [[B]]_{\mathcal{M}'}$$

holds.

(2) However, it need not be the case that

$$E'(\omega, s) \subseteq [[A \to B]]_{\mathcal{M}'} \text{ and } E'(\omega, t) \subseteq [[A]]_{\mathcal{M}'}.$$

Therefore $V_{\mathcal{M}'}(\omega, s : (A \to B)) = 1$ and $V_{\mathcal{M}'}(\omega, t : A) = 1$ need not hold and thus also $V_{\mathcal{M}'}(\omega, s \cdot_A t : B) = 1$ need not be the case anymore.

Definition 10.11 (CS-Model)**.** Let CS be a constant specification. A subset model $\mathcal{M} = (W, W_0, V_1, V_0, E)$ is called a CS-subset model or a subset model for $\mathsf{JUS}_{\mathsf{CS}}$ if for all $\omega \in W_0$ and for all $(c, A) \in \mathsf{CS}$ we have

$$E(\omega, c) \subseteq [[A]]_{\mathcal{M}}.$$

We observe that updates respect CS-subset models.

Lemma 10.12. *Let* CS *be an arbitrary constant specification and let* $\mathcal{M}$ *be a* CS*-subset model. We find that* $\mathcal{M}^C$ *is a* CS*-subset model for any formula* C*.*

10.3. Soundness

Definition 10.13 (Truth in Subset Models)**.** Let

$$\mathcal{M} = (W, W_0, V_1, V_0, E)$$

be a subset model, $\omega \in W$, and $F \in \mathsf{L}_{\mathsf{JUS}}$. We define the relation $\Vdash$ as follows:

$$\mathcal{M}, \omega \Vdash F \quad \text{iff} \quad V_{\mathcal{M}}(\omega, F) = 1.$$

Theorem 10.14 (Soundness). *Let* CS *be an arbitrary constant specification. Let* $\mathcal{M} = (W, W_0, V_1, V_0, E)$ *be a* CS*-subset model and* $\omega \in W_0$*. For each formula* $F \in \mathsf{L}_{\mathsf{JUS}}$ *we have that*

$$\mathsf{JUS}_{\mathsf{CS}} \vdash F \quad \textit{implies} \quad \mathcal{M}, \omega \Vdash F.$$

Proof. As usual by induction on the length of the derivation of F. We only show the case where F is an instance of axiom (Indep).

By induction on $[C]A$ we show that for all ω

$$\mathcal{M}^C, \omega \Vdash A \qquad \text{iff} \qquad \mathcal{M}, \omega \Vdash A.$$

We distinguish the following cases.

(1) A is an atomic proposition. Trivial.

(2) A is $\neg B$. By I.H.

(3) A is $B \to D$. By I.H.

(4) A is $t : B$. Subinduction on t:

   a) t is a variable or a constant. Easy using I.H. for B.

   b) t is a term $\mathsf{up}(D)$. By assumption, we have that $C \neq D$. Hence this case is similar to the previous case.

   c) t is a term $r \cdot_D s$. We know that $t : B$ is equivalent to

   $$r : (D \to B) \wedge s : D.$$

   Using I.H. twice, we find that

   $$\mathcal{M}^C, \omega \Vdash r : (D \to B) \quad \text{and} \quad \mathcal{M}^C, \omega \Vdash s : D$$

   if and only if

   $$\mathcal{M}, \omega \Vdash r : (D \to B) \quad \text{and} \quad \mathcal{M}, \omega \Vdash s : D.$$

   Now the claim follows immediately.

(5) A is $[D]B$. Making use of the fact that A is up-independent, this case also follows using I.H. $\square$

10.4. Basic Properties

We first show that first-order beliefs are persistent in JUS. Let F be a formula that does not contain any justification operator. We have that if

t is a justification for F, then, after any update, this will still be the case. Formally, we have the following lemma.

Lemma 10.15. *For any term t and any formulas A and C we have that if A does not contain a subformula of the form $s : B$, then*

$$t : A \;\rightarrow\; [C]t : A$$

is provable.

Proof. We proceed by induction on the complexity of t and distinguish the following cases:

(1) Case t is atomic and $t \neq \mathsf{up}(C)$. Since A does not contain any evidence terms, the claim follows immediately from axiom (Indep).

(2) Case $t = \mathsf{up}(C)$. This case is an instance of axiom (Pers).

(3) Case $t = r \cdot_B s$. From $r \cdot_B s : A$ we get by (App)

$$s : B \quad \text{and} \quad r : (B \rightarrow A).$$

By I.H. we find

$$[C]s : B \quad \text{and} \quad [C]r : (B \rightarrow A).$$

Using Lemma 10.6 we conclude $[C]r \cdot_B s : A$. □

Let us now investigate higher-order beliefs. We argue that persistence should not hold in this context. Consider the following scenario. Suppose that you are in a room together with other people. Further, suppose that no announcement has been made in that room. Therefore, it is not the case that P is believed because of an announcement. Formally, this is expressed by

$$\neg\mathsf{up}(P) : P. \tag{10.2}$$

We find that

$$\text{the fact that you are in that room} \tag{10.3}$$

justifies your belief in (10.2). Let the term r represent (10.3). Then we have

$$r : \neg\mathsf{up}(P) : P. \tag{10.4}$$

Now suppose that P is publicly announced in that room. Thus we have in the updated situation

$$\mathsf{up}(P) : P. \tag{10.5}$$

Moreover, the fact that you are in that room justifies now your belief in (10.5). Thus we have $r : \mathsf{up}(P) : P$ and hence in the original situation we have

$$[P]r : \mathsf{up}(P) : P \tag{10.6}$$

and (10.4) does no longer hold after the announcement of P.

The following lemma formally states that persistence fails for higher-oder beliefs.

Lemma 10.16. *There exist formulas $r : B$ and A such that*

$$r : B \to [A]r : B$$

is not provable.

Proof. Let B be the formula $\neg\mathsf{up}(P) : P$ and consider the subset model

$$\mathcal{M} = (W, W_0, V_1, V_0, E)$$

with

$$W := \{\omega, \upsilon\} \quad W_0 := \{\omega\} \quad V_1(\upsilon, P) = 0 \quad V_0(\omega, P) = 1$$

and

$$E(\omega, r) = \{\omega\} \quad E(\omega, \mathsf{up}(P)) = \{\omega, \upsilon\}.$$

Hence $[[P]]_{\mathcal{M}} = \{\omega\}$ and thus $E(\omega, \mathsf{up}(P)) \not\subseteq [[P]]_{\mathcal{M}}$. Since $\omega \in W_0$, this yields $V(\omega, \mathsf{up}(P) : P) = 0$. Again by $\omega \in W_0$, this implies

$$V(\omega, \neg\mathsf{up}(P) : P) = 1.$$

Therefore $E(\omega, r) \subseteq [[\neg\mathsf{up}(P) : P]]_{\mathcal{M}}$ and using $\omega \in W_0$, we get

$$\mathcal{M}, \omega \Vdash r : \neg\mathsf{up}(P) : P$$

.

Now consider the updated model $\mathcal{M}^P$. We find that

$$E^{\mathcal{M}^P}(\omega, \mathsf{up}(P)) = \{\omega\}$$

and thus $E^{\mathcal{M}^P}((\omega, \mathsf{up}(P))) \subseteq [[P]]_{\mathcal{M}^P}$. Further, using $\omega \in W_0^{\mathcal{M}^P}$ we get

$$V_{\mathcal{M}^P}(\mathsf{up}(P) : P) = 1$$

and thus $V_{\mathcal{M}^P}(\neg\mathsf{up}(P) : P) = 0$. That is $\omega \notin [[(\neg\mathsf{up}(P) : P]]_{\mathcal{M}^P}$. We have $E^{\mathcal{M}^P}(\omega, r) = \{\omega\}$ and, therefore, $E^{\mathcal{M}^P}(\omega, r) \not\subseteq [[(\neg\mathsf{up}(P) : P]]_{\mathcal{M}^P}$.

With $\omega \in W_0^{\mathcal{M}^P}$ we get $\mathcal{M}^P, \omega \not\Vdash r : \neg\mathsf{up}(P) : P$. We conclude

$$\mathcal{M}, \omega \not\Vdash [P]r : \neg\mathsf{up}(P) : P. \qquad \square$$

Next, we show that $\mathsf{JUS_{CS}}$ proves an explicit form of the Ramsey axiom

$$\Box(C \to A) \leftrightarrow [C]\Box A$$

from Dynamic Doxastic Logic.

Lemma 10.17. *Let the formula $[C]s : (C \to A)$ be up-independent. Then* $\mathsf{JUS_{CS}}$ *proves*

$$s : (C \to A) \;\leftrightarrow\; [C]s \cdot_C \mathsf{up}(C) : A. \tag{10.7}$$

Proof. First observe that by (Up), we have $[C]\mathsf{up}(C) : C$.

Further, since $[C]s : (C \to A)$ is up-independent, we find by (Indep) that

$$s : (C \to A) \leftrightarrow [C]s : (C \to A).$$

Finally we obtain (10.7) using Lemma 10.6. $\square$

Frequently, completeness of public announcement logics is established by showing that each formula with announcements is equivalent to an announcement-free formula. Unfortunately, this approach cannot be employed for $\mathsf{JUS_{CS}}$ although (10.7) provides a reduction property for certain formulas of the form $[C]t : A$. The reason is the hyperintensionality of justification logic as discussed in [9] and in Chapter 4, i.e. justification logic is not closed under substitution of equivalent formulas. Because of this, the proof by reduction cannot be carried through in $\mathsf{JUS_{CS}}$, see the discussion in [14].

Conclusion

In the first part of this thesis we have introduced new sound and complete semantics for justification logics in which justifications are interpreted as sets of worlds instead of sets of formulas. There have already been similar approaches in epistemic logics and justification logics. However, they do not provide a semantics for the standard justification logics that also allows second order justifications. We have investigated several versions of standard justification logics and compared them with each other. Contrary to other interpretations of terms, hyperintensionality does not come for free when terms are interpreted as sets of worlds. We have shown how non-normal worlds can be used to regain this important aspect of justification logic. Furthermore, we have given variants of subset models which are sound and complete with respect to $\mathsf{L}^{\star}_{\mathsf{CS}}$ even in presence of the **jd**-axiom and regardless of whether the constant specification is axiomatically appropriate or not.

In the second part we have investigated how subset models can model things like probabilistic evidence, intuition and presumptions, belief extensions and contraction. We have presented logics that model these aspects and we have established corresponding subset models. Furthermore, we have investigated alternative interpretations of the sum-operator that are not necessarily monotone. This is of special interest if we work in a context where both are relevant: the D-axiom as well as justifications that do not need to be consistent among them. We have shown how subset models can be adapted to model this by introducing a preorder on the worlds. For all the presented logics and their models we have proven soundness and model existence. Unfortunately, we do not have any completeness proofs in this second part and leave this to future research.

Nevertheless, we hope that with the results so far we have been able to develop some new aspects of justification logics that will inspire future research.

Bibliography

[1] Sergei Artemov. Operational modal logic. Technical Report MSI 95–29, Cornell University, December 1995.

[2] Sergei Artemov. Justified common knowledge. *Theor. Comput. Sci.*, 357:4–22, 07 2006.

[3] Sergei Artemov. The ontology of justifications in the logical setting. *Studia Logica*, 100, 04 2012.

[4] Sergei Artemov and Elena Nogina. Introducing justification into epistemic logic. *J. Log. Comput.*, 15:1059–1073, 12 2005.

[5] Sergei N. Artemov. Explicit provability and constructive semantics. *Bulletin of Symbolic Logic*, 7(1):1–36, March 2001.

[6] Sergei N. Artemov. The logic of justification. *Rev. Symb. Log.*, 1(4):477–513, 2008.

[7] Sergei N. Artemov. On aggregating probabilistic evidence. In *Logical Foundations of Computer Science - International Symposium, LFCS 2016, Deerfield Beach, FL, USA, January 4-7, 2016. Proceedings*, pages 27–42, 2016.

[8] Sergei [N.] Artemov and Melvin Fitting. Justification logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2012 edition, 2012.

[9] Sergei N. Artemov and Melvin Fitting. *Justification Logic: Reasoning with Reasons*. Cambridge University Press, 2019.

[10] Sergei N. Artemov and Elena Nogina. Topological semantics of justification logic. In *Computer Science - Theory and Applications, Third International Computer Science Symposium in Russia, CSR 2008, Moscow, Russia, June 7-12, 2008, Proceedings*, pages 30–39, 2008.

[11] Alexandru Baltag, Nick Bezhanishvili, Aybüke Özgün, and Sonja Smets. Justified belief and the topology of evidence. In *Logic, Language, Information, and Computation - 23rd International Workshop, WoLLIC 2016, Puebla, Mexico, August 16-19th, 2016. Proceedings*, pages 83–103, 2016.

[12] Alexandru Baltag, Bryan Renne, and Sonja Smets. The logic of justified belief, explicit knowledge, and conclusive evidence. *Annals of Pure and Applied Logic*, 165:49–81, 01 2014.

[13] Alexandru Baltag, Bryan Renne, and Sonja Smets. Revisable justified belief: Preliminary report. *CoRR*, abs/1503.08141, 2015.

[14] Samuel Bucheli, Roman Kuznets, Bryan Renne, Joshua Sack, and Thomas Studer. Justified belief change. In María De Ponte and Xabier Arrazola, editors, *LogKCA-10. Proceedings of the Second International Workshop on Logic and Philosophy of Knowledge, Communication and Action.*, page 135–155, 2010.

[15] Samuel Bucheli, Roman Kuznets, and Thomas Studer. Partial realization in dynamic justification logic. In Lev D. Beklemishev and Ruy J. G. B. de Queiroz, editors, *Logic, Language, Information and Computation - 18th International Workshop, WoLLIC 2011, Philadelphia, PA, USA, May 18-20, 2011. Proceedings*, volume 6642 of *Lecture Notes in Computer Science*, pages 35–51. Springer, 2011.

[16] Samuel Bucheli, Roman Kuznets, and Thomas Studer. Realizing public announcements by justifications. *J. Comput. Syst. Sci.*, 80(6):1046–1066, 2014.

[17] Federico L. G. Faroldi and Tudor Protopopescu. A hyperintensional logical framework for deontic reasons. *Logic Journal of the IGPL*, 27:411–433, 2019.

[18] Federico L.G. Faroldi, Meghdad Ghari, Eveline Lehmann, and Thomas Studer. Impossible and conflicting obligations in justification logic. In *proceedings of DEON 2020*, 2000.

[19] Melvin Fitting. The logic of proofs, semantically. *APAL*, 132(1):1–25, February 2005.

[20] Mark. Jago. *The impossible: An essay on hyperintensionality.* Oxford University Press, 2014.

[21] Ioannis Kokkinis, Petar Maksimović, Zoran Ognjanović, and Thomas Studer. First steps towards probabilistic justification logic. *Logic Journal of IGPL*, 23(4):662–687, 2015.

[22] Roman Kuznets and Thomas Studer. Update as evidence: Belief expansion. In Sergei N. Artëmov and Anil Nerode, editors, *Logical Foundations of Computer Science, International Symposium, LFCS 2013, San Diego, CA, USA, January 6-8, 2013. Proceedings*, volume 7734 of *Lecture Notes in Computer Science*, pages 266–279. Springer, 2013.

[23] Roman Kuznets and Thomas Studer. *Logics of Proofs and Justifications.* College Publications, 04 2019.

[24] Eveline Lehmann and Thomas Studer. Subset models for justification logic. In *Logic, Language, Information, and Computation - 26th International Workshop, WoLLIC 2019, Utrecht, The Netherlands, July 2-5, 2019, Proceedings*, pages 433–449, 2019.

[25] Eveline Lehmann and Thomas Studer. Belief expansion in subset models. In *Logical Foundations of Computer Science - International Symposium, LFCS 2020, Deerfield Beach, FL, USA, January 4-7, 2020, Proceedings*, pages 85–97, 2020.

[26] Eveline Lehmann and Thomas Studer. Exploring subset models for justification logic. submitted, 2020.

[27] I. Miethe and S. Roth. *Europas Töchter: Traditionen, Erwartungen und Strategien von Frauenbewegungen in Europa.* VS Verlag für Sozialwissenschaften, 2003.

[28] Alexey Mkrtychev. Models for the logic of proofs. In Sergei Adian and Anil Nerode, editors, *LFCS'97*, volume 1234 of *LNCS*, pages 266–275. Springer, 1997.

[29] Xavier Parent. Maximality vs. optimality in dyadic deontic logic. *J. Philosophical Logic*, 43(6):1101–1128, 2014.

[30] V. Rantala. Impossible worlds semantics and logical omniscience. *Acta Philosophica Fennica*, 35:106–115, 1982. cited By 35.

[31] Veikko Rantala. Quantified modal logic: Non-normal worlds and propositional attitudes. *Studia Logica*, 41(1):41–65, 1982.

[32] Bryan Renne. Multi-agent justification logic: Communication and evidence elimination. *Synthese*, 185:43–82, 2012.

[33] Igor Sedlár and J. Podroužek. Justification logic as dynamic epistemic logic? *Proceedings of the Second ILCLI International Workshop on Logic and Philosophy of Knowledge, Communication and Action*, pages 431–442, 01 2010.

[34] Thomas Studer. Decidability for some justification logics with negative introspection. *Journal of Symbolic Logic*, 78, 06 2013.

[35] Johan van Benthem, David Fernández Duque, and Eric Pacuit. Evidence and plausibility in neighborhood structures. *CoRR*, abs/1307.1277, 2014.

[36] Johan van Benthem, David Fernández-Duque, and Eric Pacuit. Evidence logic: A new look at neighborhood structures. *Advances in Modal Logic*, 9, 01 2014.

Index

Erklärung

gemäss Art. 18 PromR Phil.-nat. 19

Name/Vorname:	Lehmann Eveline
Matrikelnummer:	98-112-014
Studiengang:	Informatik Bachelor ☐ Master ☐ Dissertation ☒
Titel der Arbeit:	Subset Semantics for Justifications
Leiter der Arbeit:	Prof. Dr. Th. Studer

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe r des Gesetzes über die Universität vom 5. September 1996 und Artikel 69 des Universitätsstatuts vom 7. Juni 2011 zum Entzug des Doktortitels berechtigt ist.

Für die Zwecke der Begutachtung und der Überprüfung der Einhaltung der Selbständigkeitserklärung bzw. der Reglemente betreffend Plagiate erteile ich der Universität Bern das Recht, die dazu erforderlichen Personendaten zu bearbeiten und Nutzungshandlungen vorzunehmen, insbesondere die Doktorarbeit zu vervielfältigen und dauerhaft in einer Datenbank zu speichern sowie diese zur Überprüfung von Arbeiten Dritter zu verwenden oder hierzu zur Verfügung zu stellen.

....	....
Ort/Datum	Unterschrift